STARKE REPRÄSENTATION
ALLGEMEINER
STOCHASTISCHER APPROXIMATIONSVERFAHREN
MIT
ANWENDUNGEN

Dissertation,

angefertigt von

Rainer Schwabe,

vorgelegt am
Fachbereich Mathematik
der
Freien Universität Berlin
~~im November 1984~~
1985

LIBRARY

Starke Repräsentation allgemeiner stochastischer Approximationsverfahren mit Anwendungen

Dissertation,

angefertigt von

Rainer Schwabe,

vorgelegt am
Fachbereich Mathematik
der
Freien Universität Berlin
im November 1984

Diese Arbeit entstand unter der Betreuung von:

Prof. Dr. Götz Kersting (Frankfurt/Main)

Prof. Dr. Viktor Kurotschka (Berlin)

Promotionsdatum 11. 1. 85

Zusammenfassung

Ist $f: \mathbb{R} \to \mathbb{R}$ eine stetige Funktion mit genau einer Nullstelle θ, dann gibt es viele effektive Methoden zur Bestimmung von θ, falls f an jeder Stelle x exakt ausgewertet werden kann. Liegen jedoch lediglich gestörte Beobachtungen vor, können Komplikationen auftreten.

Stochastische Approximationsverfahren sind nun rekursive, auf dem Fixpunktprinzip basierende Algorithmen $X_{n+1} = X_n - A_n Y_n$, die auch beim Vorliegen von Störungen (f.s.) gegen θ konvergieren. Dabei ist $Y_n = f(X_n) + E_n$ die Beobachtung von f an der Stelle X_n, versehen mit einem Fehler E_n, und $A_n > 0$ eine geeignet gewählte (zufällige) Schrittweite.

Im Falle deterministischer Schrittweiten $A_n = n^{-1}$ konnte Ruppert (1982) eine f.s. Repräsentation von X_n durch gewichtete Fehlersummen beweisen, aus der unter geeigneten Voraussetzungen an die Störungen Invarianzprinzipien, Gesetze vom iterierten Logarithmus und asymptotische Normalität gefolgert werden können. Wir übertragen dieses Ergebnis auch auf Verfahren mit zufälligen Schrittweiten und können nachweisen, daß die dabei von uns gemachten Voraussetzungen so allgemein sind, daß sie nicht nur von Störungen, die einer Martingalbedingung genügen, sondern auch von schwach abhängigen Störungen erfüllt werden. In diesem Zusammenhang beweisen wir auch eine Marcinkiewicz-Zygmund-Ungleichung für φ-mischende Folgen. Mit den dargestellten Methoden können wir schließlich eine fast sichere Repräsentation für den von Venter (1967) vorgeschlagenen adaptiven Prozeß, dessen asymptotische Varianz minimal ist, herleiten.

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41549619_0049

Inhalt

Kapitel 1: Einleitung 1

Was ist stochastische Approximation ? 1

Der Robbins-Monro-Prozeß 3

Der Kiefer-Wolfowitz-Prozeß 6

Burkholder's A_3-Prozeß 7

Adaptive Verfahren 9

Allgemeine stochastische Approximationsverfahren . . . 11

Zur Bezeichnungsweise 15

Kapitel 2: Fast sichere Konvergenz 17

Kapitel 3: Konvergenzgeschwindigkeit 28

Kapitel 4: Eine starke Repräsentation 36

Kapitel 5: Klassische stochastische Approximation 42

Der Robbins-Monro-Prozeß 43

Der Kiefer-Wolfowitz-Prozeß 50

Burkholder's A_3-Prozeß 61

Kapitel 6: Konvergenz zufällig gewichteter Summen . . . 66

Eine Zuwachsbedingung 67

Eine Momentenbedingung 75

Kapitel 7: Der Venter-Prozeß 80

Kapitel 8: Vermischtes 100

Mehrdimensionale stochastische Approximation 100

Ein Exkurs über notwendige und hinreichende Bedingungen 104

Rekursive Schätzung 107

Anhang:

I) Hilfsmittel aus der Analysis 115

II) Einige Ergebnisse aus dem Dunstkreis fast sicherer Invarianzprinzipien 118

III) Einige interessierende Ungleichungen für mischende Folgen von Zufallsvariablen 123

Literatur: 137

Vorwort

"Es läßt sich ohne sonderlich viel Witz so schreiben, daß ein anderer sehr viel haben muß, es zu verstehen".

G.C. Lichtenberg

Die Idee der vorliegenden Arbeit entstand aus einem offenen Problem meiner Diplomarbeit, welche ich unter Anleitung von Herrn Prof. Götz Kersting anfertigte. Da es sich also um dieselbe allgemeine Problemstellung handelt, habe ich nicht davor zurückgeschreckt, Teile aus der Einleitung und einige Formulierungen in Beweisen fast wörtlich zu übernehmen. Den Vorwurf des Eigen-Plagiats ertrage ich gerne, insbesondere da ich nicht der Meinung bin, daß durch fortwährendes Umformulieren Aussagen klarer oder verständlicher werden müssen. Ein größeres Problem ist vielleicht der Umfang der Arbeit, so daß in der Fülle der Aussagen die entscheidenden Ergebnisse verschwinden. Ich stehe jedoch zu der vorliegenden Form - und andernfalls hätte ich die Arbeit auch nicht so verfaßt -, da ich meine, daß eine derartige Arbeit soweit wie möglich aus sich selbst heraus verständlich sein soll. Aber vielleicht ist diese Meinung zu tadeln ?

In erster Linie möchte ich Herrn Prof. Götz Kersting dafür danken, daß er mich über die Zeit seiner Tätigkeit in Berlin und über die Betreuung meiner Diplomarbeit hinaus mit Ratschlägen unterstützt und mich davon überzeugt hat, die vorliegenden Ergebnisse zu einer Dissertation zusammenzufassen. In diesem Zusammenhang möchte ich all den

ungenannten und auch genannten Personen meinen Dank aussprechen, die in der Entstehungszeit dieser Arbeit mir Lasten abgenommen, Probleme von mir ferngehalten oder das geleistet haben, was ich am ehesten mit "moralischer Unterstützung" bezeichnen möchte. Ich danke Herrn Prof. Viktor Kurotschka für das Interesse, das er meiner Arbeit entgegengebracht hat, und für die vielen hilfreichen Bemerkungen und Hinweise. Mein Dank gilt auch Herrn Dr. Werner Wierich für seine Hinweise sowie Frau Gundula Dietrich, Herrn Bernhard Graßhoff und Frau Ulrike Lauterbach für die neuen Aspekte, die ich aus den mit ihnen geführten Diskussionen über ihre Diplomarbeiten gewonnen habe. Insbesondere möchte ich nicht vergessen, daß die Letztgenannte einen entscheidenden Teil der Arbeit auf Inhalt und Verständlichkeit überprüfte. Ganz besonders möchte ich aber Frau Edith Bräuer dafür danken, daß sie meine Arbeit erst in die richtige Form gebracht hat. Ohne ihre Hilfe wäre diese Arbeit nicht so und nicht jetzt entstanden.

Kapitel 1: Einleitung

Was ist stochastische Approximation ?

Von praktischem Interesse ist häufig, eine Nullstelle einer stetigen Funktion f von $\mathbb{R}$ nach $\mathbb{R}$ zu finden. Weitere Probleme wie

(a) das Bestimmen eines θ mit $f(\theta) = \alpha$ für einen vorgegebenen Wert α ,

(b) das Lokalisieren eines Minimums oder Maximums einer stetig differenzierbaren Funktion f ,

lassen sich auf dieses Standardproblem zurückführen. So kann man im Fall (a) zu einer Funktion g mit $g(x) = f(x) - \alpha$ und im Fall (b) unter gewissen Voraussetzungen an die Ableitung (z.B. $f'(x) = 0$ genau dann, wenn f in x ein Extremum annimmt) zu f' übergehen.

Zur Nullstellenbestimmung gibt es eine Reihe - recht einfacher - numerischer Verfahren (vgl. Stummel und Hainer (1982), S. 29 ff). Diese Verfahren sind am allgemeinen recht gut, wenn man zu jeder vorgegebenen reellen Zahl x den exakten Funktionswert f(x) bestimmen kann. Treten jedoch bei der Beobachtung des Funktionswertes Störungen auf, so können die durch diese Methoden erzielten Ergebnisse beliebig schlecht werden. So konvergiert z.B. das Intervallschachtelungs- oder Bisektionsverfahren, das auf Vorzeichenumkehr basiert, immer; sind die Fehler jedoch derart, daß der beobachtete und der wahre Funktionswert unterschiedliches Vorzeichen besitzen können, wird die so erhaltene "Lösung" im allgemeinen keine Nullstelle sein.

Die hierbei auftretenden Störungen können auf Meßungenauigkeiten zurückzuführen sein oder aber auf die Tatsache, daß die Funktion mit den zur Verfügung stehenden Mitteln nicht gemessen werden kann, wie

z.B. in der Chemie die Abhängigkeit der Reaktionsgeschwindigkeit von der Temperatur. Ein extremes und klassisches Beispiel ist das Problem, das α-Quantil einer Zufallsvariablen X mit stetiger Verteilungsfunktion F , d.h. θ mit $F(\theta) = \alpha$, zu bestimmen (vgl. Wasan (1969), S. 8 ff): Sei X z.B. die tödliche Dosis eines Insektizids, so daß $X(\omega)$ die jeweilige kritische tödliche Menge für ein Insekt ω aus der Gesamtmenge Ω aller Insekten ist. Wendet man nun die Menge x dieses Insektizids auf ein Insekt ω an, so kann man nicht $F(x)$ beobachten, sondern lediglich Erfolg (1) oder Mißerfolg (0), je nach dem, ob $X(\omega) \leq x$ oder $X(\omega) > x$ gilt, d.h. ob die Dosis x für das behandelte Insekt ω tödlich ist oder nicht. *)

Unter stochastischen Approximationsverfahren versteht man nun solche iterativen (Fixpunkt-)Verfahren $X_{n+1} = X_n - A_n Y_n$, wobei A_n noch zu spezifizieren und Y_n der "beobachtete" Funktionswert an der

---

*) In diesem Beispiel kann man "natürlich" Insektizid durch Tränengas und Insekten durch Demonstranten ersetzen. X bezeichne dann wieder die für eine gewisse Wirkung benötigte Dosis, wobei vom ökonomischen Standpunkt als Wirkung Handlungsunfähigkeit und α möglichst groß gewählt wird. Vom "humanitären" Standpunkt - sofern man in diesem Zusammenhang davon sprechen kann - sollte bei zu untersuchender Wirkung "körperliche Schäden" α möglichst klein gewählt werden. Glücklicherweise werden die hier weiter beschriebenen Methoden nicht für solch menschenverachtendes Denken und Handeln benutzt werden, sei es, daß es weniger zeitaufwendige und damit effektivere Verfahren gibt (vgl. hierzu auch die Fußnote auf Seite 5), oder sei es - was eher zu befürchten ist - daß solche Mittel eingesetzt werden, ohne daß ihre Wirkungen bedacht werden.

Stelle X_n zum Zeitpunkt n ist, für die X_n auch beim Vorliegen von Störungen in den Beobachtungen Y_n gegen eine Nullstelle konvergiert. Auf Grund der Zufälligkeit der Fehler handelt es sich bei den X_n um Zufallsvariablen, und man betrachtet daher im allgemeinen fast sichere Konvergenz.

Im folgenden werden wir zuerst klassische stochastische Approximationsverfahren vorstellen, und am Ende dieses Kapitels einen Überblick über unsere Ergebnisse geben.

Der Robbins-Monro-Prozeß

Sei $M: \mathbb{R} \to \mathbb{R}$ und $\theta \in \mathbb{R}$ mit $M(x)(x - \theta) > 0$ für alle $x \neq \theta$. Hier und im folgenden seien stets alle Funktionen meßbar vorausgesetzt. Falls M stetig ist, ist θ (einzige) Nullstelle von M , und M ist wachsend im Punkte θ .

Ein einfacher auf dem Fixpunktprinzip beruhender Iterationsalgorithmus zum Auffinden von θ ist:

$$x_{n+1} = x_n - a_n M(x_n)$$

mit $a_n > 0$, $n \in \mathbb{N}$, und Startwert x_1 . Treten bei der Beobachtung des Funktionswertes an der Stelle x_n zum Zeitpunkt n zufällige Störungen U_n auf, so daß man lediglich $Y_n = M(x_n) + U_n$ erhält, und beachtet man noch, daß somit Zufallsvariable X_n vorliegen, so liegt es nahe, folgenden Prozeß zu betrachten:

$$X_{n+1} = X_n - a_n(M(X_n) + U_n) .$$

Um mit diesem Verfahren den Median (d.h. das 0,5-Quantil) einer Verteilung bestimmen zu können, ist es notwendig,

$$\lim_{n\to\infty} a_n = 0 \quad \text{und} \quad \Sigma_{n=1}^{\infty} a_n = \infty$$

zu fordern. Denn:

Sei F die zugrunde liegende Verteilungsfunktion. Dann ist $M = F - \frac{1}{2}$ zu wählen, und die Beobachtungen $Y_n = M(X_n) + U_n$ können nur die Werte $+\frac{1}{2}$ (Erfolg) und $-\frac{1}{2}$ (Mißerfolg) annehmen. Gilt nun $\lim_{n\to\infty} a_n = 0$ nicht, so kann $(X_n)_{n\in\mathbb{N}}$ nicht konvergieren. Gilt andererseits $\Sigma_{n=1}^{\infty} a_n = A < \infty$, dann konvergiert $(X_n)_{n\in\mathbb{N}}$, jedoch bei "unglücklicher" Wahl des Startwertes $-|X_1 - \theta| > \frac{1}{2} A$ - sicherlich nicht gegen θ.

Ein klassisches Beispiel für solche a_n ist die Folge $(n^{-1})_{n\in\mathbb{N}}$.

Wir können nun einen relativ einfachen, rekursiven Algorithmus zum Schätzen von θ angeben:

Seien X_1, U_n, $n \in \mathbb{N}$, Zufallsvariable.
Die Folge $(X_n)_{n\in\mathbb{N}}$, definiert durch

$$X_{n+1} = X_n - an^{-1}Y_n$$

mit

$$Y_n = M(X_n) + U_n,$$

$a > 0$, heißt Robbins-Monro-Prozeß.

Die ersten Ergebnisse für diesen Prozeß wurden unter recht rigiden Voraussetzungen an die Fehler U_n bewiesen. Dazu wurde davon ausgegangen, daß die Fehler U_n bei gegebenem $X_n = x$ nur von x abhängen, insbesondere also unabhängig von den vorhergehenden Beobachtungen sind, und $E(U_n) = 0$ gilt. Das heißt mit anderen Worten:
Zu jedem $x \in \mathbb{R}$ existiert eine Verteilung F_x mit Erwartung 0, und für

alle $n \in \mathbb{N}$ gilt: Die bedingte Verteilung von U_n gegeben $X_n = x$ ist F_x. Solche Fehler werden wir in Übereinstimmung mit der üblichen Definition bedingt unabhängig nennen (vgl. Loève (1978) II, S. 290).

Unter gewissen zusätzlichen Regularitätsbedingungen konnten Robbins und Monro (1951) quadratische Konvergenz und damit Konvergenz in Wahrscheinlichkeit für den von ihnen erfundenen Prozeß zeigen. Blum (1954a) gelang der Nachweis fast sicherer Konvergenz. Chung (1954) folgerte asymptotische Normalität von $n^{\frac{1}{2}}(X_n - \theta)$ unter der Voraussetzung konstanter Varianz c^2 für alle F_x (die Fehler bilden dann einen schwach stationären Prozeß) und konnte zeigen, daß die Wahl von $a_n = an^{-1}$ in dem Sinne optimal ist, daß sich für $a_n = an^{-\rho}$, $\frac{1}{2} < \rho < 1$, schlechtere Konvergenzgeschwindigkeiten ergeben. Gesetze vom iterierten Logarithmus finden sich bei Major (1973), Heyde (1974) sowie Gaposhkin und Krasulina (1974). McLeish (1976) bewies ein Verteilungs-Invarianzprinzip, und mit Kersting's (1977) Repräsentation des Robbins-Monro-Prozesses durch eine gewichtete Summe von unabhängigen Zufallsvariablen ist es möglich, ein fast sicheres Invarianzprinzip zu zeigen (vgl. Heyde (1981)), aus dem sich die meisten der vorhergehenden Ergebnisse folgern lassen. *)

---

*) Trotz seiner guten asymptotischen Eigenschaften scheint der Robbins-Monro-Prozeß zur Quantilschätzung nicht praktikabel, wenn α weit von 0,5 entfernt ist. Sei z.B. $\alpha = 0{,}9$ und der Startwert X_1 sei das gesuchte θ. Mit Wahrscheinlichkeit 0,1 wird man einen Mißerfolg beobachten, und $x_2 = \theta - 0{,}9a$ setzen. Mit Hilfe der Integralabschätzung $\sum_{i=2}^{n} \frac{1}{i} \leq \int_1^n \frac{1}{x}\,dx$ ergibt sich sofort, daß in diesem Fall $|x_n - \theta| \leq 0{,}1a$ frühestens nach $n = 2\,981 \tilde{\geq} e^8$ und $x_n \leq \theta$ frühestens nach $n = 8\,103 \tilde{\geq} e^9$ Iterationsschritten gelten kann ! (Vgl. Wetherill (1975). S. 173 f).

Für viele Ergebnisse konnte die Forderung der bedingten Unabhängigkeit der Fehler durch eine schwächere Martingaleigenschaft ($E(U_{n+1} \mid X_1, U_1, \ldots, U_n) = 0$ f.s.) ersetzt und Chung's (1954) Voraussetzung der konstanten Varianzen erheblich abgeschwächt werden.

Der Kiefer-Wolfowitz-Prozeß

Sei $M: \mathbb{R} \to \mathbb{R}$ stetig und $\theta \in \mathbb{R}$ mit $\liminf_{h \to 0} \left(\frac{M(x+h) - M(x)}{h}(x - \theta)\right) >$ für alle $x \neq \theta$. Das heißt insbesondere: θ ist Minimum von M und einziges lokales Extremum.

Setzt man M stetig differenzierbar voraus, so erscheint es sinnvoll, zum Auffinden von θ den weiter oben beschriebenen Iterationsalgorithmus $x_{n+1} = x_n - a_n M'(x_n)$ zu benutzen. Häufig kann jedoch M' nicht ausgewertet werden, und man muß $M'(x_n)$ durch eine Schätzung, z.B. $(2c_n)^{-1}(M(x_n + c_n) - M(x_n - c_n))$, ersetzen. Hierdurch schleicht sich ein systematischer Fehler b_n ein, für den möglichst $\lim_{n \to \infty} b_n = 0$ gelten sollte. Für $b_n = (2c_n)^{-1}(M(x_n + c_n) - M(x_n - c_n)) - M'(x_n) = M'(\xi_n) - M'(x_n)$, für ein ξ_n zwischen $x_n - c_n$ und $x_n + c_n$, aus unserem Beispiel läßt sich das durch $\lim_{n \to \infty} c_n = 0$ erreichen.

Treten bei der Beobachtung der Funktionswerte an den Stellen $x_n + c_n$ und $x_n - c_n$ zum Zeitpunkt n zusätzlich zufällige Störungen $U_n = U_n^+ - U_n^-$ auf, so daß man lediglich $Y_n^+ = M(x_n + c_n) + U_n^+$ bzw. $Y_n^- = M(x_n - c_n) + U_n^-$ erhält, und beachtet man wieder, daß Zufallsvariablen vorliegen, so liegt es nahe, folgenden Prozeß zu betrachten:

Seien X_1 , U_n , $n \in \mathbb{N}$, Zufallsvariable.

Die Folge $(X_n)_{n \in \mathbb{N}}$, definiert durch

$$X_{n+1} = X_n - an^{-1}Y_n$$

mit

$$Y_n = (2c_n)^{-1}(M(X_n + c_n) - M(X_n - c_n) + U_n) ,$$

$a > 0$, $c_n = cn^{-\gamma}$, $c > 0$, $0 < \gamma < \frac{1}{2}$, heißt Kiefer-Wolfowitz-Prozeß.

Unter der Voraussetzung bedingt unabhängiger Fehler, das heißt hier: Die Verteilung von U_n^+ und U_n^- gegeben $X_n = x$ ist F_{x+c_n} bzw. F_{x-c_n} , und unter gewissen Regularitätsvoraussetzungen konnten Kiefer und Wolfowitz (1952) für diesen von ihnen erfunden Prozeß Konvergenz in Wahrscheinlichkeit nachweisen. Blum (1954a) zeigte auch für diesen Prozeß fast sichere Konvergenz, und Burkholder (1956) und Sacks (1958) bewiesen asymptotische Normalität. Gesetze vom iterierten Logarithmus finden sich wieder bei Heyde (1974) sowie Gaposhkin und Krasulina (1974).

Burkholder's A_3-Prozeß

Sei $M: \mathbb{R} \to \mathbb{R}$ zweimal stetig differenzierbar und $\theta \in \mathbb{R}$ mit $M'(x)(x - \theta) > 0$ für alle $x \neq \theta$. Die Funktion M ist somit konkav auf $(-\infty, \theta)$, konvex auf (θ, ∞) und besitzt in θ einen Wendepunkt.

Das Problem, einen Wendepunkt einer Funktion zu finden, scheint auf den ersten Blick von rein theoretischem Interesse zu sein. Möchte man aber den Modalwert einer stetig differenzierbaren Dichte bestimmen, dann kann man weder die Dichte noch deren Ableitung auswerten, sondern lediglich die zugehörige Verteilungsfunktion F (Erfolg/Mißerfolg). Das führt zu der vorgegebenen Problemstellung mit $M = - F$.

Im Iterationsalgorithmus $x_{n+1} = x_n - a_n M''(x_n)$ muß man wieder $M''(x_n)$ durch eine Schätzung, z.B. $c_n^{-2}(M(x_n + c_n) + M(x_n - c_n) - 2M(x_n))$, ersetzen. Für den hierbei auftretenden systematischen Fehler b_n läßt sich $\lim_{n\to\infty} b_n = 0$ wegen

$$b_n = c_n^{-2}(M(x_n + c_n) + M(x_n - c_n) - 2M(x_n)) - M''(x_n)$$

$$= \frac{1}{2} c_n^{-2}(M''(\xi_n^+) + M''(\xi_n^-)) - M''(x_n)$$

nach Taylor, für gewisse ξ_n^+, ξ_n^- zwischen $x_n - c_n$ und $x_n + c_n$, wieder durch $\lim_{n\to\infty} c_n = 0$ erreichen.

Treten bei der Beobachtung der Funktionswerte an den Stellen x_n, $x_n + c_n$ und $x_n - c_n$ zum Zeitpunkt n zusätzlich zufällige Störungen $U_n = U_n^+ + U_n^- - 2U_n^0$ auf, so daß man lediglich $Y_n^0 = M(x_n) + U_n^0$, $Y_n^+ = M(x_n + c_n) + U_n^+$ bzw. $Y_n^- = M(x_n - c_n) + U_n^-$ erhält, so gelangt man zu folgendem Prozeß:

Seien X_1, U_n, $n \in \mathbb{N}$ Zufallsvariable.
Die Folge $(X_n)_{n\in\mathbb{N}}$, definiert durch

$$X_{n+1} = X_n - an^{-1}Y_n$$

mit

$$Y_n = c_n^{-2}(M(X_n + c_n) + M(X_n - c_n) - 2M(X_n) + U_n) \quad ,$$

$a > 0$, $c_n = cn^{-\gamma}$, $c > 0$, $0 < \gamma < \frac{1}{4}$, nennen wir $\underline{A_3\text{-Prozeß}}$.

Burkholder (1956) konnte für den von ihm vorgeschlagenen Prozeß fast sichere Konvergenz und asymptotische Normalität nachweisen.

Adaptive Verfahren

Ausgehend von Robbins-Monro-Prozeß mit bedingt unabhängigen Fehlern stellt sich die Frage: Kann man die Konvergenzeigenschaften durch geeignete Wahl von a verbessern ?

Die Antwort ist: Ja ! Setzt man zusätzlich voraus, daß $\lambda := M'(\theta)$ existiert und daß die Varianz von F_x für $x \to \theta$ gegen die Varianz σ^2 von F_θ strebt, so erhält man aus Ergebnissen von Chung (1954) und Sacks (1958), daß $n^{\frac{1}{2}}(X_n - \theta)$ asymptotisch normalverteilt mit Erwartung 0 und Varianz $a^2\sigma^2(2\lambda a - 1)^{-1}$ ist, wenn nur $2\lambda a > 1$ gilt (für $2\lambda a \leq 1$ konnten Major und Révész (1973) eine schlechtere Konvergenzrate als $n^{-\frac{1}{2}}$ nachweisen). Wählt man nun $a = \lambda^{-1}$, so erhält man minimale Varianz der Grenzverteilung, somit optimales Konvergenzverhalten.

In den meisten Anwendungen ist jedoch λ nicht bekannt. Es liegt daher nahe, zum Vergleich n für a einen aus den bisherigen Beobachtungen gewonnenen Schätzwert für λ^{-1} zu wählen. Solche stochastische Approximationsprozesse, bei denen der Faktor vor Y_n keine festen reellen Zahlen, sondern von den bisherigen Beobachtungen abhängige Zufallsvariablen sind, nennen wir adaptive Prozesse.

Als Beispiel betrachten wir einen von Venter (1967) vorgeschlagenen und von Fabian (1968) gründlich modifizierten Approximationsprozeß, den wir im weiteren Venter-Fabian nennen werden.

Seien X_1 , U_n , V_n , $n \in \mathbb{N}$ Zufallsvariable.

X_n , $n \geq 2$, sei definiert durch:

$$X_{n+1} = X_n - A_n Y_n$$

mit

$$Y_n = \frac{1}{2}(M(X_n + c_n) - M(X_n - c_n)) + U_n \quad ,$$

$$D_n = n^{-1} \sum_{i=1}^{n} (2c_i)^{-1} (M(X_i + c_i) - M(X_i - c_i)) \quad ,$$

$$c_n = cn^{-\gamma} \quad , \quad c > 0 \, , \; 0 < \gamma < \frac{1}{2} \quad ,$$

und

$$A_n = \begin{cases} (n \log n)^{-1} & \text{für } D_n > \log n \\ n^{-1+\alpha} & \text{für } D_n < n^{-\alpha} \\ (nD_n)^{-1} & \text{sonst} \end{cases}$$

für $n \geq 3$, $A_1 = A_2 = 1$, $0 < \alpha < \frac{1}{2}$.

Die Vorgehensweise ist folgende: Man beobachtet M nicht an der Stelle $X_n = \vartheta_n$, sondern ähnlich wie beim Kiefer-Wolfowitz-Prozeß an den Stellen $X_n + c_n$ und $X_n - c_n$ und erhält $Y_n^+ = M(X_n + c_n) + U_n^+$ bzw. $Y_n^- = M(X_n - c_n) + U_n^-$. $\frac{1}{2}(Y_n^+ + Y_n^-)$ ist eine Schätzung für $M(X_n)$ und $(2c_n)^{-1}(Y_n^+ - Y_n^-)$ eine Schätzung für λ . Letztere wird durch Mittelung über alle bisherigen Schätzungen verbessert, und man erhält D_n . Aus technischen Gründen muß D_n^{-1} beschnitten werden, was schließlich zu A_n führt. Die Fehler U_n und V_n ergeben sich aus $U_n = \frac{1}{2}(U_n^+ + U_n^-)$ bzw. $V_n = U_n^+ - U_n^-$.

Venter (1967) und Fabian (1968) konnten unter gewissen Regularitätsvoraussetzungen nachweisen, daß $(X_n)_{n \in \mathbb{N}}$ fast sicher konvergiert und daß $n^{\frac{1}{2}}(X_n - \vartheta)$ asymptotisch normalverteilt mit Erwartung 0 und Varianz $\frac{1}{2}c^2\lambda^{-2}$ ist. Die minimal mögliche Varianz beim Robbins-Monro-Prozeß war

$\sigma^2\lambda^{-2}$, jedoch läßt sich der Faktor $\frac{1}{2}$ dadurch erklären, daß beim Venter-Prozeß zu jedem Zeitpunkt doppelt so viele Beobachtungen wie beim Robbins-Monro-Prozeß gemacht werden, so daß man bezogen auf die Anzahl der Beobachtungen gleiche Güte erhält.

Faßt man D_n nicht als Schätzer für $M'(\theta)$, sondern als Schätzer für $M'(X_n)$ auf, dann kann sich der Venter-Prozeß auch über das bekanntermaßen recht gute Newton-Verfahren der Numerik motivieren.

Zum Kiefer-Wolfowitz-Prozeß läßt sich im allgemeinen kein "bestes" adaptives Verfahren konstruieren, da die optimale Wahl von γ vom qualitativen – und damit nicht schätzbaren – Verhalten der zugrundeliegenden Funktion abhängt:

Ist M dreimal stetig differenzierbar in einer Umgebung von θ und $M'''(\theta) \neq 0$, dann wird die optimale Konvergenzrate $n^{-\frac{1}{3}}$ mit $\gamma = \frac{1}{6}$ erreicht (vgl. Fabian (1971)). Ist hingegen $M'''(\theta) = 0$ und M fünfmal stetig differenzierbar in einer Umgebung von θ, so lassen sich bessere Konvergenzraten für $\frac{1}{10} \leq \gamma < \frac{1}{6}$ erzielen (vgl. Sacks (1958)). Gilt zusätzlich $M^{(5)}(\theta) \neq 0$, so können wir für $\gamma = \frac{1}{10}$ die optimale Konvergenzrate $n^{-\frac{2}{5}}$ herleiten, und im Fall $M^{(5)}(\theta) = 0$ läßt sich die Konvergenz eventuell noch verbessern, wobei $\gamma < \frac{1}{10}$ zu wählen ist usw. (vgl. Kapitel 5).

Allgemeine stochastische Approximationsverfahren

Bislang wurde vorausgesetzt, daß die auftretenden Beobachtungsfehler bedingt unabhängig sind oder zumindest eine Martingaleigenschaft besitzen. (Ein neuerer Überblick über die in diesem Fall bekannten Ergebnisse findet sich bei Nevel'son und Has'minskii (1973) und Schmetterer (1979)).

In den praktischen Anwendungen lassen sich solche idealen Versuchsbedingungen jedoch selten verwirklichen, und es ist mit Auftreten von Abhängigkeiten zu rechnen. Es ist aber zu hoffen, daß die Konvergenzeigenschaften von stochastischen Approximationsmethoden erhalten bleiben, sofern die Abhängigkeiten nicht zu stark werden.

Sei wieder $f: \mathbb{R} \to \mathbb{R}$, $\theta \in \mathbb{R}$ mit $f(x)(x - \theta) > 0$ für alle $x \neq \theta$. Wir beschränken uns auf - recht allgemeine - stochastische Approximationsalgorithmen $X_{n+1} = X_n - A_n Y_n$ zur Bestimmung von θ , bei denen sich die Beobachtung Y_n aus dem Funktionswert $f(X_n)$, einer zufälligen Störung E_n und einem systematischen Fehler B_n zusammensetzen kann:

$$X_{n+1} = X_n - A_n(f(X_n) + E_n + B_n) .$$

Um auch adaptive Prozesse behandeln zu können, lassen wir für die Schrittweiten A_n positive Zufallsvariablen zu, die die beim Robbins-Monro-Prozeß motivierten Eigenschaften $\lim_{n\to\infty} A_n = 0$ und $\sum_{n=1}^{\infty} A_n = \infty$ zumindest fast sicher besitzen. Weiterhin ist es sinnvoll, $\lim_{n\to\infty} B_n = 0$ f.s. zu fordern, um den Einfluß der systematischen Fehler zu dämpfen.

Wie man leicht nachprüfen kann, ist die hier betrachtete Approximationsformel allgemein genug, um die klassischen Methoden (Robbins-Monro-Prozeß; Kiefer-Wolfowitz-Prozeß für differenzierbare Funktionen M) und auch adaptive Prozesse, wie den Venter-Prozeß, zu umfassen (vgl. Kapitel 5 und 7).

In Kapitel 2 werden wir unter gewissen Regularitätsvoraussetzungen an die Funktion f und die systematischen Fehler B_n fast sichere Konvergenz von X_n gegen θ beweisen, sofern nur $\sum_{n=1}^{\infty} A_n E_n$ fast sicher konvergiert. Die zufälligen Fehler E_n wurden ein derartiges Resultat erstmals explizit in einer Arbeit von Chung (1978) genügt, die auch mit N-dimensionaler

stochastischer Approximation, d.h. Bestimmen einer Nullstelle einer stetigen Funktion von $\mathbb{R}^N$ nach $\mathbb{R}^N$, befaßt (vgl. Kapitel 8). Da wir uns auf den eindimensionalen Fall beschränken wollen, können wir, einer Beweisidee von Blum (1954a) folgend, fast sichere Konvergenz nicht nur erheblich schneller, sondern auch unter schwächeren Voraussetzungen als Ljung (1978) nachweisen.

Unter weiteren Voraussetzungen sowohl an die Funktion f, als auch an die Schrittweiten A_n, die systematischen Fehler B_n und die zufälligen Störungen E_n (geringere Abhängigkeit) werden wir in Kapitel 3 eine Aussage über die Konvergenzgeschwindigkeit und in Kapitel 4 eine fast sichere Repräsentation von X_n durch eine Summe gewichteter Störungen beweisen. Das von Ruppert (1982) für deterministische Schrittweiten $A_n = n^{-1}$ formulierte Ergebnis, werden wir dabei in zwei Richtungen verallgemeinern: Einerseits - und das ist das Wichtigere - gelingt es uns, auch für adaptive Prozesse eine Repräsentation zu finden, andererseits ist diese auch im deterministischen Fall für eine größere Klasse stochastischer Approximationsprozesse sinnvoll anwendbar.

Diese Anwendung auf klassische stochastische Approximationsverfahren werden wir in Kapitel 5 durchführen, und unter der zusätzlichen Annahme, daß die Störungen in den Beobachtungen einem fast sicheren Invarianzprinzip genügen, erhalten wir konkretere Ergebnisse wie Invarianzprinzipien und asymptotische Normalität.

Im Kapitel 7 werden wir eine starke Repräsentation für den adaptiven Venter-Prozeß formulieren, aber bevor dieses sinnvoll geschehen kann, müssen wir uns in Kapitel 6 davon überzeugen, daß unsere Voraussetzung der Konvergenz gewisser gewichteter Fehler-

summen nicht zu speziell gewählt ist. Zuerst bemerken wir, daß diese Voraussetzung unter einer Meßbarkeitsforderung an die Schrittweiten A_n erfüllt ist, wenn nur die Störungen E_n neben gleichmäßig beschränkten Varianzen die Martingaleigenschaft $E(E_{n+1}|X_1,E_1,\ldots,E_n) = 0$ f.s. besitzen. Damit lassen sich die Ergebnisse von Venter (1967) und Fabian (1968) beweisen. Hauptinhaltlich des Kapitels 6 ist es jedoch, für eine andere große Klasse von Störungen, die einer recht komplizierten Zuwachsbedingung genügen, unter zusätzlichen Voraussetzungen an die Schrittweiten A_n die Konvergenz der gewichteten Fehlersummen nachzuweisen. Diese Zuwachsbedingung läßt sich durch eine schärfere aber leichter verständliche Bedingung an die absoluten Momente von Fehlersummen ersetzen, und wir werden im dritten Teil des Anhangs zeigen, daß diese Bedingung an die Momente von einer weiten Klasse von schwach abhängigen Fehlerprozessen erfüllt wird. Als Nebenergebnis fällt dabei eine Ungleichung vom Marcinkiewicz-Zygmund-Typ für φ-mischende Störungen ab.

Das letzte Kapitel beinhaltet einige Bemerkungen über N-dimensionale stochastische Approximation, über notwendige und hinreichende Bedingungen, sowie über rekursive Schätzung.

Zur Bezeichnungsweise

Für $x \in \mathbb{R}$ bezeichnen wir mit $[x]$ die größte ganze Zahl k mit $k \leq x$. Insbesondere gilt also $[x] \leq x < [x] + 1$.

Für $f,g: \mathbb{R} \to \mathbb{R}$ und $\xi \in \mathbb{R} \cup \{-\infty, \infty\}$ mit $g > 0$ in einer Umgebung von ξ schreiben wir $f(x) = O(g(x))$ für $x \to \xi$, falls $\limsup_{x \to \xi} |f(x)|g(x)^{-1} < \infty$ gilt, und $f(x) = o(g(x))$ für $x \to \xi$, falls $\lim_{x \to \xi} f(x)g(x)^{-1} = 0$ gilt. Im Falle $\xi = \infty$ verzichten wir auf das Anhängsel "für $x \to \infty$". Offensichtlich impliziert die o-Aussage die O-Aussage. Für Folgen $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ reeller Zahlen, $(b_n)_{n \in \mathbb{N}}$ nicht negativ, definieren wir das Landau-Symbol O für Folgen und schreiben $a_n = O(b_n)$, falls ein $C \in \mathbb{R}$ existiert, so daß $|a_n| \leq Cb_n$ für alle $n \in \mathbb{N}$ gilt. Faßt man Folgen als Funktionen eingeschränkt auf $\mathbb{N}$ auf und betrachtet $\xi = \infty$, so sind für positive $(b_n)_{n \in \mathbb{N}}$ beide Definitionen äquivalent, wobei wir das offensichtliche Suffix "für $n \to \infty$" unterdrückt haben. Das Landau-Symbol o für Folgen: $a_n = o(b_n)$ bedeutet entsprechend $\lim_{n \to \infty} a_n b_n^{-1} = 0$. Für Folgen $(X_n)_{n \in \mathbb{N}}$, $(Y_n)_{n \in \mathbb{N}}$ von Zufallsvariablen, $Y_n \geq 0$, schreiben wir $X_n = O(Y_n)$ f.s. bzw. $X_n = o(Y_n)$ f.s., falls für fast alle ω aus dem zugrunde liegenden Raum Ω die Beziehung $X_n(\omega) = O(Y_n(\omega))$ bzw. $X_n(\omega) = o(Y_n(\omega))$ gilt. Entsprechend sind $X(t) = O(Y(t))$ f.s. bzw. $X(t) = o(Y(t))$ f.s. für stochastische Prozesse X,Y definiert. Weiterhin soll $f(x) = g(x) + O(h(x))$ bedeuten: $f(x) - g(x) = O(h(x))$ etc.

Schließlich sei noch bemerkt, daß wir mit $\xrightarrow{\mathcal{D}}$ die Konvergenz in Verteilung bezeichnen, und $\xrightarrow{\mathcal{D}}$ mit dem Anhängsel in $D[0,1]$ steht für

die schwache Konvergenz von zufälligen Funktionen im Raum $\mathcal{D}[0,1]$ der rechtsseitig stetigen reellwertigen Funktionen auf dem Intervall $[0,1]$, deren linksseitige Grenzwerte existieren, versehen mit der Skorohod-Topologie (vgl. Billingsley (1968), S. 109 ff). Entsprechend ist der Raum $\mathcal{D}[0,\infty)$ definiert, wobei die Funktionen dabei auf dem halboffenen Intervall $[0,\infty)$ operieren.

Kapitel 2: Fast sichere Konvergenz

In diesem Kapitel werden wir fast sichere Konvergenz recht allgemeiner stochastischer Approximationsverfahren $X_{n+1} = X_n - A_n Y_n$ untersuchen. Dabei werden wir für die Schrittweiten A_n Zufallsvariablen zulassen. Die Beobachtung Y_n setzt sich aus dem Funktionswert $f(X_n)$ der zu untersuchenden Funktion f , einem systematischen Fehler B_n und einer zufälligen Störung E_n zusammen.

Wir betrachten also stochastische Approximationsalgorithmen der Form

$$X_{n+1} = X_n - A_n(f(X_n) + E_n + B_n) \quad . \tag{2.1}$$

Für diesen Prozeß konnte Ljung (1978) fast sichere Konvergenz nachweisen, wobei er die Bedingungen an die Störungen auf die entscheidende Forderung der fast sicheren Konvergenz von $\Sigma_{n=1}^{\infty} A_n E_n$ reduzierte. Da er jedoch N-dimensionale stochastische Approximation behandelte, traten beweistechnische Komplikationen auf, die zusätzliche Regularitätsbedingungen an die Funktion f erforderten. Wir folgen direkt dem Beweis von Blum (1954a) und können zeigen:

(2.2) Satz

Genügen $f: \mathbb{R} \to \mathbb{R}$, $\theta \in \mathbb{R}$, Zufallsvariablen X_1, A_n, E_n, B_n mit $A_n > 0$ und der durch

$$X_{n+1} = X_n - A_n(f(X_n) + E_n + B_n)$$

definierte Prozeß den Voraussetzungen:

(A) $f(x)(x - \theta) > 0$ für alle $x \neq \theta$,

(B) $\inf_{\delta_1 \leq |x-\theta| \leq \delta_2} |f(x)| > 0$ für alle $\delta_2 \geq \delta_1 > 0$,

(C) es ex. $\alpha, \beta > 0$ mit $|f(x)| \leq \alpha + \beta |x - \theta|$ für alle x ,

(D) $A_n \to 0$, $\Sigma_{n=1}^{\infty} A_n = \infty$ f.s. ,

(E) $\Sigma_{n=1}^{\infty} A_n E_n$ konvergiert f.s. ,

(F) $B_n \to 0$ f.s. ,

(G) es ex. $K > 0$ mit $\liminf_{n\to\infty} \frac{B_n}{f(X_n)} 1_{\{|X_n|>K\}} > -1$ f.s. ,

dann gilt:

$$X_n \to \theta \text{ f.s.}$$

Bemerkungen:

i) Voraussetzung (A) besagt lediglich, daß $f(x) > 0$ für $x > \theta$ und $f(x) < 0$ für $x < \theta$ gilt. Ist zusätzlich f stetig, dann folgt aus (A), daß θ Nullstelle, und zwar einzige Nullstelle, von f ist. Unter (A) ist somit für stetiges f Voraussetzung (B) immer erfüllt. Als Beispiel einer Funktion, die (A), aber nicht (B) genügt, möge man sich eine geeignete Sägezahnkurve vorstellen: etwa f mit $f(x) = x - [x] - 1$ für $x < 0$, $f(0) = 0$ und $f(x) = -f(-x)$ für $x > 0$.

ii) Nach Voraussetzung (C) darf $|f|$ für $|x| \to \infty$ höchstens linear wachsen, was etwa beim Vorliegen von Lipschitz-Stetigkeit gesichert ist. Daß (C) bereits im deterministischen Fall benötigt wird, möge folgendes Beispiel illustrieren: Sei $A_n = n^{-1}$, $B_n = E_n = 0$, und f

mit $f(x) = x(1 + |x|^{\varepsilon})$ für ein $\varepsilon > 0$. Sei $c := 2^{\frac{1}{\varepsilon}} + 1$ und für x_1 gelte $|x_1| \geq c^{\frac{1}{\varepsilon}}$. Dann folgt induktiv $|x_n| \geq (cn)^{\frac{1}{\varepsilon}}$, denn:

$$|x_{n+1}| = |x_n|\,|1 - \frac{1}{n}(1 + |x_n|^{\varepsilon})| \geq (cn)^{\frac{1}{\varepsilon}}(c - 1) = (2cn)^{\frac{1}{\varepsilon}} \geq (c(n+1))^{\frac{1}{\varepsilon}}.$$

iii) Sei $A_n = an^{-1}$, $a > 0$, deterministisch, und seien E_n, $n \in \mathbb{N}$, unabhängig mit $E(E_n) = 0$ und $\mathrm{Var}(E_n) \leq c^2 < \infty$. Wegen $\sum_{n=1}^{\infty} \mathrm{Var}(A_n E_n) < \infty$ ist dann bereits Voraussetzung (E) erfüllt (vgl. Loève (1977) I, S. 248).

iv) Die Voraussetzung (G) an die systematischen Fehler ist sehr implizit, ist aber sicherlich erfüllt, wenn keine systematischen Fehler auftreten, d.h. $B_n = 0$ für alle n. Verstärkt man (B) zu

(B') $\inf_{|x-\theta| \geq \delta} |f(x)| > 0$ für alle $\delta > 0$,

dann folgt (G) wegen $|B_n f(X_n)^{-1} 1_{\{|X_n| > K\}}| \leq |B_n| (\inf_{|x-\theta| \geq K} |f(x)|)^{-1}$ bereits aus (F). Jedoch schließt (B') Funktionen, die für $x \to \infty$ bzw. $x \to -\infty$ gegen 0 streben, z.B. f mit $f(x) = xe^{-x^2}$, aus. In Satz (2.4) werden wir für unsere Anwendungszwecke angemessenere Voraussetzungen an die systematischen Fehler formulieren.

Beweis:

Ohne Beschränkung der Allgemeinheit (O.B.d.A.) sei $\theta = 0$. Daß dieses tatsächlich keine Einschränkung ist, werden wir hier exemplarisch auch für die anderen Beweise zeigen. Dazu führen wir eine Koordinatentransformation durch:

$$Y_n := X_n - \theta, \; n \in \mathbb{N}, \qquad g(x) := f(x + \theta), \; x \in \mathbb{R}.$$

g erfüllt die Voraussetzungen des Satzes mit $\theta = 0$, und

$$Y_{n+1} = X_{n+1} - \theta$$

$$= X_n - \theta - A_n(f(X_n) + E_n + B_n)$$

$$= Y_n - A_n(g(Y_n) + E_n + B_n)$$

ist von der Form (2.1), so daß wegen $X_n \to \theta$ f.s. genau dann, wenn $Y_n \to 0$ f.s. bereits alles gezeigt ist.

Wir wenden uns nun dem eigentlichen Beweis zu. Dabei wird unsere Vorgehensweise hier und auch in späteren Beweisen so sein, daß wir pfadweise die Behauptung rein analytisch aus den Voraussetzungen herleiten:

Sei (Ω,A,P) der zugrunde liegende Wahrscheinlichkeitsraum und sei

$$\Omega' := \{\omega | \omega \in \Omega \ , \ \lim_{n\to\infty} A_n(\omega) = 0 \ , \ \Sigma_{n=1}^{\infty} A_n(\omega) = \infty \ ,$$

$$\lim_{n\to\infty} B_n(\omega) = 0 \ , \ \liminf_{n\to\infty} B_n(\omega) f(X_n(\omega))^{-1} \ 1_{\{|X_n|>K\}}(\omega) > -$$

$$\Sigma_{n=1}^{\infty} A_n(\omega) E_n(\omega) \text{ konvergiert}\} \ .$$

Auf Grund der Voraussetzungen (D) bis (G) gilt $\Omega' \in A$ und $P(\Omega') = 1$. Es genügt somit, $\lim_{n\to\infty} X_n(\omega) = 0$ für alle $\omega \in \Omega'$ zu zeigen.

Sei also $\omega \in \Omega'$. Die Realisationen $X_n(\omega)$, $A_n(\omega)$ usw. werden wir wie vereinbart mit x_n , a_n usw. bezeichnen, und wir werden uns im weiteren direkt auf die Voraussetzungen (D) bis (G) beziehen, wenn wir eigentlich die Eigenschaft für unsere spezielle Realisation und

somit ohne den Zusatz "f.s." meinen. Wir bemerken noch, daß sämtliche auftretenden "Konstanten" von ω abhängen können.

Der Beweis wird nun in drei Schritten durchgeführt:

(1) $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$,

(2) $|\lim_{n\to\infty} x_n| < \infty$,

(3) $\lim_{n\to\infty} x_n = 0$,

wobei wir zu jedem einzelnen Schritt einen Widerspruchsbeweis durchführen werden:

Zu (1):

Annahme: $\liminf_{n\to\infty} x_n < \limsup_{n\to\infty} x_n$!

Fall 1: $\limsup_{n\to\infty} x_n > 0$.

Dann existieren $a,b > 0$ mit

(4) $\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n$.

Das Intervall $[a,b]$ wird somit von der Folge $(x_n)_{n\in\mathbb{N}}$ beliebig oft überquert. Unser Ziel wird es sein, zu zeigen, daß ab einem genügend großen Zeitpunkt N eine derartige Überquerung nicht mehr stattfinden kann:

Zu $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ existiert nach Voraussetzung (D), (E) bzw. (F) ein $N = N(\varepsilon_1,\varepsilon_2,\varepsilon_3) \in \mathbb{N}$, so daß für alle $m,n \geq N$ mit $n > m$

(5) $a_m \leq \varepsilon_1$,

(6) $|\sum_{i=m}^{n-1} a_i e_i| \leq \varepsilon_2$,

(7) $\quad |b_m| \leq \varepsilon_3$

gilt. Dabei ist (6) gerade die Cauchy-Eigenschaft einer konvergenten Reihe. Aus beweistechnischen Gründen wählen wir

(5') $\quad \varepsilon_1 := \min(\frac{b-a}{4\alpha}, \frac{1}{4\beta}) > 0$,

(6') $\quad \varepsilon_2 := \frac{b-a}{4} > 0$,

(7') $\quad \varepsilon_3 := \min(\alpha, \inf\limits_{a \leq x \leq b} f(x)) > 0$.

Wegen (4) existieren $m,n \geq N$, $n > m$, mit

(8) $\quad x_n > b$, $x_m < a$ und $a \leq x_i \leq b$ für $m < i < n$.

Für diese m,n folgt nun durch Iteration aus unserer Rekursionsformel (2.1):

$$x_n - x_m = - \Sigma_{i=m}^{n-1} a_i (f(x_i) + b_i) - \Sigma_{i=m}^{n-1} a_i e_i \ .$$

Nach (8) und Wahl von ε_3 gilt $f(x_i) + b_i \geq 0$ für $m < i < n$.
Unter Verwendung von (5) bis (7) erhalten wir die Abschätzung

$$x_n - x_m \leq - a_m f(x_m) - a_m b_m - \Sigma_{i=m}^{n-1} a_i e_i \leq - a_m f(x_m) + \varepsilon_1 \varepsilon_3 + \varepsilon_2$$

$$\leq - a_m f(x_m) + \frac{1}{2}(b-a) \ .$$

Für $x_m \geq 0$ ergibt sich wegen $f(x_m) \geq 0$ sofort ein Widerspruch zu (8). Betrachten wir also den verbleibenden Fall $x_m < 0$.
Dann gilt nach Voraussetzung (C):

$$- f(x_m) = |f(x_m)| \leq \alpha + \beta |x_m| = \alpha - \beta x_m \ ,$$

folglich mit (5) und $x_n > 0$:

$$x_n - x_m \leq a_m(\alpha - \beta x_m + \beta x_n) + \frac{1}{2}(b - a)$$

$$\leq \frac{3}{4}(b - a) + \frac{1}{4}(x_n - x_m) ,$$

was wiederum einen Widerspruch zu $x_n - x_m > b - a$ (vgl. (8)) liefert. Damit ist Fall 1 abgehandelt.

Fall 2: $\limsup_{n\to\infty} x_n \leq 0$.

Dann gilt insbesondere $\liminf_{n\to\infty} x_n < 0$.
Nach Übergang von x_n zu $y_n = - x_n$ und von f zu g mit $g(x) = - f(- x)$ liegt wieder Fall 1 und damit ein Widerspruch vor.

Somit ist unsere Annahme falsch, also Behauptung (1) richtig.

Zu (2):

Annahme: $\lim_{n\to\infty} x_n = \infty$!

Dann existiert ein N_o , so daß $x_n > K$ für alle $n \geq N_o$ gilt. Nach der merkwürdigen Voraussetzung (G), die nur hier eingeht, gibt es sogar ein $N \geq N_o$ mit $f(x_n) + b_n \geq 0$ für alle $n \geq N$. Nach Iteration erhalten wir für $n \geq N$ aus unserer Rekursionsformel (2.1) die Abschätzung:

$$x_{n+1} = x_N - \Sigma_{i=N}^{n} a_i(f(x_i) + b_i) - \Sigma_{i=N}^{n} a_i e_i$$

$$\leq x_N + \sup_{m\geq N} |\Sigma_{i=N}^{m} a_i e_i| .$$

Da der letzte Term nach (E) eine endliche Konstante ist, erfolgt Widerspruch. Die andere

Annahme: $\lim_{n\to\infty} x_n = - \infty$!

kann wieder nach dem zu (1) beschriebenen Übergang widerlegt werden.

Zu (3):

<u>Annahme:</u> $\lim_{n\to\infty} x_n > 0$!

Dann gibt es $\delta_1, \delta_2 > 0$ mit $\delta_1 < \lim_{n\to\infty} x_n < \delta_2$.
Wir definieren uns eine Hilfsgröße $c := \frac{1}{2} \inf_{\delta_1 \leq x \leq \delta_2} f(x)$, welche nach (B) größer als 0 ist. Unter Berücksichtigung von (F) existiert ein N , so daß $\delta_1 \leq x_n \leq \delta_2$ und $|b_n| \leq c$ für $n \geq N$ erfüllt ist. Für $n \geq N$ gilt also insbesondere

$$f(x_n) + b_n \geq c \; ,$$

und Iteration der Rekursionsformel liefert:

$$x_{n+1} = x_N - \Sigma_{i=N}^{n} a_i (f(x_i) + b_i) - \Sigma_{i=N}^{n} a_i e_i$$

$$\leq x_N - c \, \Sigma_{i=N}^{n} a_i + \sup_{m \geq N} |\Sigma_{i=N}^{m} a_i e_i| \; .$$

Wiederum ist nach (E) der letzte Term eine endliche Konstante, und aus der zweiten Voraussetzung unter (D) folgt $\lim_{n\to\infty} x_n = -\infty$ und somit ein Widerspruch zur Annahme. Die andere

<u>Annahme:</u> $\lim_{n\to\infty} x_n < 0$!

führt wieder aus Symmetriegründen zum Widerspruch. □

Die impliziten Voraussetzungen an die systematischen Fehler B_n sind (außer [illegible] Bemerkung iii) zum [illegible] aufgeführten Fällen, [illegible] schwer nachzuprüfen. [illegible] wissen wir [illegible] die Gestalt der B_r :

Es gibt Funktionen $f_n : \mathbb{R} \to \mathbb{R}$, $n \in \mathbb{N}$, mit

$$B_n = f_n(X_n) - f(X_n) .$$

Daher erscheint es sinnvoll, stochastische Approximationsalgorithmen der Form

$$X_{n+1} = X_n - A_n(f_n(X_n) + E_n) \tag{2.3}$$

zu betrachten und Forderungen an die Funktionen f_n statt an die systematischen Fehler B_n zu stellen. Derartige Rekursionsformeln, die in der einschlägigen Literatur oft als dynamische stochastische Approximation bezeichnet werden, wurden bereits von Dupac (1966) untersucht. Wir wollen uns hier jedoch nicht mit der Allgemeinheit dieser Theorie auseinandersetzen, sondern folgende für die von uns betrachteten speziellen Prozesse wichtige Aussage beweisen:

(2.4) Satz:

Genügen $\theta \in \mathbb{R}$, Funktionen $f_n : \mathbb{R} \to \mathbb{R}$ und Zufallsvariablen X_1, A_n, E_n mit $A_n > 0$ den Voraussetzungen:

(A) für alle $\varepsilon > 0$ ex. ein $N(\varepsilon)$, so daß $f_n(x)(x - \theta) > 0$ für alle $n \geq N(\varepsilon)$ und $|x - \theta| \geq \varepsilon$,

(B) $\liminf_{n\to\infty} \inf_{\delta_1 \leq |x-\theta| \leq \delta_2} |f_n(x)| > 0$ für alle $\delta_2 \geq \delta_1 > 0$,

(C) es ex. $\alpha, \beta > 0$ und N_0 mit $|f_n(x)| \leq \alpha + \beta|x - \theta|$ für alle $n \geq N_0$ und alle x ,

(D) $A_n \to 0$, $\sum_{n=1}^{\infty} A_n = \infty$ f.s. ,

(E) $\sum_{n=1}^{\infty} A_n E_n$ konvergiert f.s. ,

dann gilt für den durch

$$X_{n+1} = X_n - A_n(f_n(X_n) + E_n)$$

definierten Prozeß:

$$X_n \to \theta \quad \text{f.s.}$$

Bemerkung:

(A) bis (C) stellen naheliegende Abschwächungen von Dupac's (1966) Voraussetzungen dar. Da wir keine Konvergenz von f_n gegen eine geeignete Funktion f fordern, lassen sich keine Zusammenhänge zwischen (A) bis (C) und (2.2) (F) und (G) herleiten. Tatsächlich muß selbst bei punktweiser Konvergenz von f_n gegen f der systematische Fehler $B_n = f_n(X_n) - f(X_n)$ nicht gegen 0 streben. Daher haben wir die Aussagen der Sätze (2.2) und (2.4) gesondert formuliert. Der

Beweis

ist jedoch lediglich eine Variation des gleichen Themas. Insbesondere betrachten wir Realisationen und o.B.d.A sei $\theta = 0$.

Zu (2.2) (1):

Für $b > a > 0$ sichern uns (D) und (E) die Existenz eines $N \geq \max(N_0, N(a))$ mit $N(a)$ aus (A), so daß $a_m \leq \varepsilon_1$ und $|\Sigma_{i=m}^{n-1} a_i e_i| \leq \varepsilon_2$ mit $\varepsilon_1 = \min(\frac{b-a}{4(\alpha + a\beta)}, \frac{1}{2\beta})$ bzw. $\varepsilon_2 = \frac{b-a}{4}$ für alle $n > m \geq N$ erfüllt ist. Unter der Annahme

$$\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n$$

mit $a > 0$ gibt es zu diesem N aber $n > m \geq N$, so daß

$$x_n > b \,,\; x_m < a \quad \text{und} \quad a \leq x_i \leq b \quad \text{für } m < i < n$$

gilt. Die Abschätzung

$$x_n - x_m = -\sum_{i=m}^{n-1} a_i f_i(x_i) - \sum_{i=m}^{n-1} a_i e_i$$

$$\leq -a_m f_m(x_m) + \varepsilon_2$$

führt für $x_m \geq 0$ über

$$x_n - x_m \leq a_m(\alpha + \beta x_m) + \varepsilon_2 \leq a_m(\alpha + \beta a) + \varepsilon_2 \leq \tfrac{1}{2}(b - a)$$

und für $x_m < 0$ über

$$x_n - x_m \leq a_m(\alpha - \beta x_m + \beta x_m) + \varepsilon_2 \leq \tfrac{1}{2}(b - a) + \tfrac{1}{2}(b - x_m)$$

zum Widerspruch.

Zu (2.2) (2):

Die Annahme "$\lim_{n\to\infty} x_n = \infty$!" impliziert die Existenz von $N \geq N(1)$ mit $x_n \geq 1$ und $f_n(x_n) \geq 0$ für alle $n \geq N$. Die daraus resultierende Ungleichung

$$\sup_{n\geq N} x_n \leq x_N + \sup_{n\geq N} \left|\sum_{i=N}^{n-1} a_i e_i\right| < \infty$$

widerlegt dieses.

Zu (2.2) (3):

Für $\delta_2 > \delta_1 > 0$ existiert nach (A) und (B) ein $N' \geq N(\delta_1)$ und ein $c > 0$, so daß $\inf_{\delta_1 \leq x \leq \delta_2} f_n(x) \geq c$ für alle $n \geq N'$ erfüllt ist.

Unter der Annahme "$\delta_1 < \lim_{n\to\infty} x_n < \delta_2$!" gibt es nun ein $N \geq N'$ mit $\delta_1 < x_n < \delta_2$ für $n \geq N$, und die Abschätzung

$$x_{n+1} \leq x_N - c\sum_{i=N}^{n} a_i + \sup_{n\geq N} \left|\sum_{i=N}^{n} a_i e_i\right|$$

liefert den gewünschten Widerspruch.

Kapitel 3: Konvergenzgeschwindigkeit

Nachdem die fast sichere Konvergenz unserer stochastischen Approximationsverfahren gesichert ist, möchten wir weitere Aussagen über das Verhalten von X_n für $n \to \infty$ machen. Das im vorangegangenen Kapitel verwendete Beweisverfahren ist nun trotz seiner Einfachheit von einer derartigen Güte, daß sich unter naheliegenden Zusatzvoraussetzungen für die Konvergenzgeschwindigkeit $X_n - \theta = O(n^{-\delta})$ für gewisse $\delta > 0$ ergibt. Da wir an die zufälligen Störungen und an die systematischen Fehler dieselben Bedingungen stellen werden, gehen wir zu allgemeinen Fehlern $E_n' := E_n + B_n$ (bzw. $E_n' := E_n + f_n(X_n) - f(X_n)$) über und betrachten Algorithmen der Form

$$X_{n+1} = X_n - A_n(f(X_n) + E_n') \quad . \tag{3.1}$$

(3.2) Satz:

Genügen $f: \mathbb{R} \to \mathbb{R}$, $\theta \in \mathbb{R}$, Zufallsvariablen X_1, A_n, E_n' mit $A_n > 0$, sowie der durch

$$X_{n+1} = X_n - A_n(f(X_n) + E_n')$$

definierte Prozeß und δ mit $0 < \delta < 1$ den Voraussetzungen:

(A) f ist beschränkt in einer Umgebung von θ,

(B) $\liminf_{x \to \theta} \frac{f(x)}{x - \theta} = K > 0$,

(C) $X_n \to \theta$ f.s.,

(D) $n^{\delta}A_n \to 0$ f.s. ,

(E) $\sum_{n=1}^{\infty} n^{\delta}A_n E'_n$ konvergiert f.s. ,

(F) $\liminf_{n\to\infty} nA_n > \frac{\delta}{K}$ f.s. ,

dann gilt:

$$n^{\delta}(X_n - \theta) \to 0 \quad \text{f.s.}$$

Bemerkungen:

i) Die Formulierung des Satzes ist angelehnt an Ruppert's (1982) Lemma 3.1, jedoch ist unser Beweis für den eindimensionalen Fall sehr viel übersichtlicher. (Insbesondere erscheint mir die Anwendung eines Lemmas von Derman und Sacks (1959) bei Ruppert (1982), S. 180, 4. Zeile von unten, nicht möglich. Jedoch läßt sich die benötigte Aussage relativ leicht nachweisen, und ich werde sie im Anhang als Lemma (A. 3) aufführen.)

ii) Ist f in θ differenzierbar, dann ist (B) gleichbedeutend mit $f'(\theta) > 0$. Allgemein besagt diese Voraussetzung lediglich, daß f in einer Umgebung von θ durch eine Gerade mit positiver Steigung von 0 wegbeschränkt ist. Man kann (B) unter der Nebenbedingung $f(x)(x - \theta) > 0$ für alle $x \neq \theta$, durch die äquivalente Forderung

(B') es gibt ein $K' > 0$ und $\varepsilon > 0$ mit $|f(x)| \geq K'|x - \theta|$ für alle $|x - \theta| < \varepsilon$

ersetzen. Aus (B') folgt trivialerweise (B) mit $K \geq K'$. Umgekehrt läßt sich zeigen, daß für jedes $K' < K$ ein $\varepsilon = \varepsilon(K') > 0$ existiert, so daß (B') er-

füllt ist. Insbesondere ist also auch γ die kleinste obere Schranke der nach (B') zulässigen K' .

Wollen wir $K = \infty$ zulassen, um auch in θ "steile" Funktionen wie $\operatorname{sgn} := 1_{(0,\infty)} - 1_{(-\infty,0)}$ oder f mit $f(x) = \sqrt{|x|}\,\operatorname{sgn}(x)$ betrachten zu können, muß (F) entsprechend in $\liminf_{n\to\infty} nA_n > 0$ abgeändert werden.

iii) Dank (C) können wir hier auf die die Konvergenz gegen θ sichernde Voraussetzungen des letzten Kapitels über das Verhalten von f außerhalb einer Umgebung von θ verzichten. In der Anwendung werden wir sie jedoch zum Nachweis dieser impliziten Forderung benötigen.

iv) Nach (F) dürfen die Schrittweiten A_n nicht schneller als n^{-1} gegen 0 streben. Insbesondere muß nicht nur $\Sigma_{n=1}^{\infty} A_n$ divergieren, sondern zusätzlich sind Folgen der Gestalt $((n\log n)^{-1})_{n\in\mathbb{N}}$ ausgeschlossen. Eine naive Betrachtungsweise der Voraussetzungen (D) und (E) läßt außerdem hoffen, optimale Konvergenzrate zu erhalten, falls auch $\limsup_{n\to\infty} nA_n < \infty$ gilt. Aus (D) und (F) ergibt sich, daß $\delta < 1$ keine zusätzliche Einschränkung darstellt, jedoch erlauben wir uns hier den Luxus redundanter Information.

v) Seien E_n' , $n \in \mathbb{N}$, unabhängig identisch verteilt mit $E(E_n') = 0$ und $\operatorname{Var}(E_n') < \infty$. Ist $A_n = an^{-1}$, deterministisch, dann ist (E) nur für alle $\delta < \frac{1}{2}$ erfüllt; im allgemeinen ist also nicht mit einer Konvergenzrate von $n^{-\frac{1}{2}}$ oder besser rechnen. Ist $A_n = an^{-\alpha}$, $\frac{1}{2} < \alpha < 1$, dann können wir Konvergenz nur für $\delta < \alpha - \frac{1}{2}$ nachweisen; in der Tat ist die Konvergenzrate schlechter als im Fall $\alpha = 1$ (vgl. Chung (1954)).

vi) Es sei noch am Rande bemerkt, daß mit Hilfe des Abel'schen Konvergenzkriteriums für Reihen (vgl. Knopp (1974), S. 238) aus (Ⓔ) die fast sichere Konvergenz von $\sum_{n=1}^{\infty} n^{\delta'} A_n b'_n$ für alle $\delta' < \delta$ folgt.

Beweis:

O.B.d.A. können wir $\theta = 0$ setzen.

Wir werden zeigen:

(1) $\quad (n-1)^{\delta} X_n \to 0$ f.s. ,

woraus wegen

$$|n^{\delta} X_n| \leq (2(n-1))^{\delta} |X_n| = 2^{\delta} |(n-1)^{\delta} X_n| ,$$

für $n \geq 2$, die Behauptung folgt.

Sei Ω' wieder die Teilmenge des zugrunde liegenden Wahrscheinlichkeitsraumes, auf der die Voraussetzungen punktweise erfüllt sind, also $P(\Omega') = 1$, und wir betrachten Realisationen. Sei also $\omega \in \Omega'$. Es ist $(n-1)^{\delta} X_n = (n-1)^{\delta} X_n(\omega) \to 0$ zu zeigen.

Zuerst wollen wir gemäß einer Idee von Ruppert (1982) die Formel (3.1) zu einer Rekursion für $(n-1)^{\delta} x_n$ umschreiben:

$$\begin{aligned} n^{\delta} x_{n+1} &= n^{\delta}(x_n - a_n(f(x_n) + e'_n)) \\ &= (n-1)^{\delta} x_n + (n^{\delta} - (n-1)^{\delta}) x_n - n^{\delta} a_n f(x_n) - n^{\delta} a_n e'_n . \end{aligned}$$

Um nun den zweiten Term etwas näher zu untersuchen, entwickeln wir g mit $g(x) = x^{\delta}$ nach Taylor um n und erhalten:

$$(n-1)^{\delta} = n^{\delta} - \delta n^{\delta-1} + O(n^{\delta-2}) .$$

Nach (C) ist $x_n = O(1)$, folglich:

$$(n^\delta - (n-1)^\delta)x_n = \delta n^{\delta-1}x_n + O(n^{\delta-2})$$

Als endgültige Rekursionsformel ergibt sich somit:

$$n^\delta x_{n+1} = (n-1)^\delta x_n + n^{\delta-1}(\delta x_n - na_n f(x_n)) + d_n$$

mit

$$d_n = -n^\delta a_n e'_n + O(n^{\delta-2}) \ .$$

Iterieren liefert für $n > m$:

$$(2) \qquad (n-1)^\delta x_n = (m-1)^\delta x_m + \Sigma_{i=m}^{n-1} i^{\delta-1}(\delta x_i - ia_i f(x_i)) + \Sigma_{i=m}^{n-1} d_i \ .$$

Da nach Voraussetzung (E) und wegen $\delta < 1$

$$(3) \qquad \Sigma_{n=1}^{\infty} d_n \text{ konvergiert ,}$$

werden die d_n keine Auswirkung auf die Konvergenz bzw. Divergenz von $(n-1)^\delta x_n$ haben.

Als nächstes wenden wir uns den Voraussetzungen (B) und (F) zu. Nach (F) existieren K' , $0 < K' < K$, und $\rho > 0$ mit

$$\liminf_{n\to\infty} na_n > \frac{\delta + \rho}{K'} \ .$$

Wie bereits bemerkt, gibt es nach (B) zu diesem K' ein $\varepsilon' > 0$, so daß $x^{-1}f(x) \geq K'$ für $x \neq 0$ mit $|x| < \varepsilon'$ gilt. Somit existiert unter Berücksichtigung von (C) ein $N_o = N_o(\rho, K', \varepsilon')$, so daß

$$(4) \qquad K'na_n \geq \delta + \rho \ ,$$

$$(5) \qquad |x_n| < \varepsilon'$$

[illegible] ist. Insbesondere [illegible] für $x_n > 0$, [illegible] $\geq N_o$, aus (5)

$$f(x_n) \geq K'x_n \ ,$$

folglich mit (4):

(6) $$\delta x_n - n a_n f(x_n) \leq \delta x_n - K' n a_n x_n \leq - \rho x_n$$

eine Abschätzung für einen in (2) auftretenden Term.

Der Rest des Beweises der Behauptung (1) zerfällt in zwei Schritte:

(7) $$\liminf_{n\to\infty} (n-1)^{\delta} x_n \leq 0 \leq \limsup_{n\to\infty} (n-1)^{\delta} x_n ,$$

(8) $$\liminf_{n\to\infty} (n-1)^{\delta} x_n = 0 = \limsup_{n\to\infty} (n-1)^{\delta} x_n ,$$

und wird wieder im Widerspruchsverfahren durchgeführt.

Zu (7):

Unter der

Annahme: $\liminf_{n\to\infty} (n-1)^{\delta} x_n > 0$!

existiert $c > 0$ und $N \geq N_0$ mit

(9) $$n^{\delta} x_n \geq (n-1)^{\delta} x_n \geq c$$

für alle $n \geq N$. Insbesondere gilt für $n \geq N$ also $x_n > 0$ und wir können unsere Abschätzung (6) in Verbindung mit (9) in die iterierte Rekursionsformel (2) einsetzen:

$$(n-1)^{\delta} x_n \leq (N-1)^{\delta} x_N - \rho \Sigma_{i=N}^{n-1} i^{\delta-1} x_i + \Sigma_{i=N}^{n-1} d_i$$

$$\leq (N-1)^{\delta} x_N - \rho c \Sigma_{i=N}^{n-1} i^{-1} + \Sigma_{i=N}^{n-1} d_i .$$

Unter Berücksichtigung der Konvergenz von $\Sigma_{i=N}^{\infty} d_i$ ergibt sich

$$\lim_{n\to\infty} (n-1)^{\delta} x_n = -\infty$$

im Widerspruch zur Annahme.

Der zweite Teil der Behauptung ($\limsup_{n\to\infty}(n-1)^{\delta}x_n \geq 0$) folgt wieder sofort nach Übergang von x_n zu $y_n = -x_n$ und von f zu g mit $g(x) = -f(-x)$.

Zu (8):

Nach Voraussetzung (A) existiert ein $\varepsilon_o > 0$ und ein $c < \infty$ mit $|f(x)| \leq c$ für $|x| < \varepsilon_o$. Weiterhin gibt es wegen (C), (D) und (3) zu ε mit $0 < \varepsilon < \varepsilon_o$ ein $N \geq N_o$, so daß für alle $m,n \geq N$, $m < n$, gilt:

(10) $\quad |x_m| < \varepsilon$,

(11) $\quad m^{\delta}a_m < \varepsilon$,

(12) $\quad |\Sigma_{i=m}^{n-1} d_i| < \varepsilon$.

Annahme: $\limsup_{n\to\infty}(n-1)^{\delta}x_n > 0$!

Dann können wir unser N bestimmendes $\varepsilon > 0$ so wählen, daß $(3 + c)\varepsilon < \limsup_{n\to\infty}(n-1)^{\delta}x_n$ gilt, und wegen $\liminf_{n\to\infty}(n-1)^{\delta}x_n \leq 0$ nach (7) gibt es $n > m \geq N$ mit

$$(n-1)^{\delta}x_n > (3 + c)\varepsilon \text{ , } (m-1)^{\delta}x_m < \varepsilon \text{ und}$$

$$\varepsilon \leq (i-1)^{\delta}x_i \leq (3 + c)\varepsilon \text{ für } m < i < n .$$

Eine derartige Überquerung des Intervalls $[\varepsilon,(3 + c)\varepsilon]$ der Länge $(2 + c)\varepsilon$ ist aber nicht mehr möglich, denn (10) bis (12) in die iterierte Rekursionsformel (2) eingesetzt, liefert unter Beachtung von $\varepsilon < \varepsilon_o$ und von (6) wegen $N \geq N_o$ und $x_i \geq 0$, $m < i < n$, die Abschätzung:

$$(n-1)^{\delta}x_n - (m-1)^{\delta}x_m = \sum_{i=m}^{n-1} i^{\delta-1}(\delta x_i - i a_i f(x_i)) + \sum_{i=m}^{n-1} a_i$$

$$\leq m^{\delta-1}\delta x_m - m^{\delta} a_m f(x_m) + \varepsilon$$

$$< \varepsilon + \varepsilon c + \varepsilon ,$$

womit die Annahme widerlegt ist.

Nach den üblichen Übergängen folgt wieder der andere Teil der Behauptung ($\liminf_{n\to\infty}(n-1)^{\delta}x_n = 0$) aus Symmetriegründen. □

Kapitel 4: Eine starke Repräsentation

Ausgehend von den Ergebnissen des vorhergehenden Kapitels drängt sich uns nun die Frage nach der optimalen Konvergenzrate auf. Wie verhält sich zum Beispiel $n^{\Delta}(X_n - \theta)$ für ein $\Delta > 0$, falls für alle $\delta < \Delta$ die Konvergenzrate $X_n - \theta = O(n^{-\delta})$ f.s. vermittels der f.s. Konvergenz von $\Sigma_{n=1}^{\infty} n^{\delta} A_n E_n'$ gesichert ist. Im Falle deterministischer Schrittweiten $A_n = an^{-1}$, $a > 0$, konnte Ruppert (1982) eine fast sichere Repräsentation von $n^{\Delta}(X_n - \theta)$ durch gewichtete Fehlersummen beweisen. Eine erste Annäherung dieser Art für den Robbins-Monro-Prozeß mit bedingt unabhängigen Fehlern durch eine gewichtete Summe unabhängiger, identisch verteilter Zufallsvariablen findet sich bei Kersting (1977). Wir können aber Ruppert's (1982) Aussage auch auf den Fall ausweiten, daß sich die Schrittweiten nur ungefähr wie an^{-1} , $a > 0$, verhalten:

(4.1) Satz

i) Genügen $f: \mathbb{R} \to \mathbb{R}$, $\theta \in \mathbb{R}$, Zufallsvariablen X_1, A_n, E_n' mit $A_n > 0$, $n \geq 1$, sowie der durch

$$X_{n+1} = X_n - A_n(f(X_n) + E_n')$$

definierte Prozeß, $a > 0$ und Δ mit $0 < \Delta \leq 1$ den Voraussetzungen:

(A) f ist differenzierbar in θ mit $f'(\theta) = \lambda > 0$, und es ex. $\eta > 0$ mit $f(x) = \lambda(x - \theta) + O(|x - \theta|^{1+\eta})$ für $x \to \theta$,

(B) $X_n \to \theta$ f.s.,

(C) es ex. $\rho > 0$ mit $nA_n = a + O(n^{-\rho})$ f.s.,

(D) $\Sigma_{n=1}^{\infty} n^{\delta} A_n E_n'$ konvergiert f.s. für alle $\delta < \Delta$,

dann gilt:

(1) $$n^{\Delta}(X_{n+1} - \theta) = - n^{-\lambda a+\Delta} \Sigma_{i=1}^{n} i^{\lambda a} A_i E_i' + O(n^{-\varepsilon}) \text{ f.s.}$$

für ein $\varepsilon > 0$.

ii) Ist anstelle von (E)

(E_1) $\lambda a = \Delta$ erfüllt, dann existiert eine Zufallsvariable Z , so daß

(2) $$n^{\lambda a}(X_{n+1} - \theta) = - \Sigma_{i=1}^{n} i^{\lambda a} A_i E_i' + Z + O(1) \text{ f.s.}$$

gilt.

Bemerkungen:

i) Im Falle deterministischer Schrittweiten $A_n = an^{-1}$ erscheint die vorliegende Formulierung auf den ersten Blick sehr viel spezieller als Ruppert's (1982) Ergebnis. Jedoch können sich die E_n' wieder aus systematischen Fehlern B_n und zufälligen Störungen E_n zusammensetzen: $E_n' = E_n + B_n$. Nach geeigneter Identifikation besitzen Ruppert's (1982) Voraussetzungen an die Fehlerterme die Gestalt:

(D_1) es ex. $b \in \mathbb{R}$ und $\beta > 0$ mit $n^{\Delta} B_n = b + O(n^{-\beta})$ f.s.,

(D_2) $\Sigma_{n=1}^{\infty} n^{-1+\delta} E_n$ konvergiert f.s. für alle $\delta < \Delta$,

woraus die Voraussetzung (D) folgt, die somit genügend allgemein ist. Weiterhin sind wir in der Wahl von Δ nicht so eingeschränkt, so daß wir bereits für klassische stochastische Approximationsprozesse zusätzliche Ergebnisse formulieren können (vgl. Kapitel 5).

ii) Nach (A) ist θ Nullstelle von f . Bereits der zweite Teil dieser Voraussetzung alleine impliziert die Differenzierbarkeit von f in θ mit $f'(\theta) = \lambda$, jedoch garantiert selbst stetige Differenzierbarkeit von f

in θ nicht die Existenz eines geeigneten $\eta > 0$:

Als Beispiel betrachten wir f mit $f(x) = x + x(\log|x|)^{-1}$ für $|x| \le c := e^{-2}$, $x \neq 0$, $f(0) = 0$ und $f(x) = \frac{1}{4}(x + c\frac{x}{|x|})$ für $|x| > c$. Es ist $\theta = 0$. f ist stetig differenzierbar mit $f'(0) = 1$. Unter der Annahme, daß f die Voraussetzung für ein η erfüllte, müßte $x^{-1}f(x) - \lambda = (\log|x|)^{-1} = O(|x|^{\eta})$ für $x \to 0$ gelten, was aber für alle $\eta > 0$ falsch ist.

Hinreichend hingegen ist zweimalige Differenzierbarkeit von f in einer Umgebung U von θ und $\sup_{x\in U}|f''(x)| < \infty$:

Für x aus dieser Umgebung gilt nach Taylor-Entwicklung $f(x) = f(\theta) + f'(\theta)(x - \theta) + O(|x - \theta|^2)$ für $x \to \theta$, was wegen $f(\theta) = 0$ die Behauptung mit $\eta = 1$ liefert.

iii) (C) gilt trivialerweise für $A_n = an^{-1}$, $a > 0$.

iv) Für $\delta < \Delta$ sind unter (A) bis (E) bzw. (E_1) die Voraussetzungen des Satzes (3.2) mit $K = \lambda$ erfüllt, so daß bereits $X_n - \theta = O(n^{-\delta})$ f.s. gilt.

Beweis:

O.B.d.A. können wir wieder $\theta = 0$ setzen. Nach (A) und (B) gilt:

$$f(X_n) = \lambda X_n + O(|X_n|^{1+\eta}) \text{ f.s.},$$

folglich mit (C):

$$A_n f(X_n) = an^{-1}\lambda X_n + R_n$$

mit

$$R_n = O(n^{-1}|X_n|^{1+\eta} + n^{-1-\rho}|X_n|) \text{ f.s.}$$

Somit erhalten wir für unseren Prozeß die Rekursionsformel:

$$X_{n+1} = X_n - an^{-1}\lambda X_n - R_n - A_n E_n' .$$

Wir können uns nun des Terms $an^{-1}\lambda X_n$ entledigen, indem wir - ähnlich wie im Beweis von (3.2) - unsere Gleichung mit $n^{\lambda a}$ multiplizieren und zu einer Rekursion für $(n-1)^{\lambda a}X_n$ übergehen. Dazu entwickeln wir g mit $g(x) = x^{\lambda a}$ um n nach Taylor:

$$(n-1)^{\lambda a} = n^{\lambda a} - \lambda a n^{\lambda a-1} + \nu_n$$

mit

$$\nu_n = O(n^{\lambda a-2}) .$$

Damit läßt sich unsere Rekursion wie folgt:

$$n^{\lambda a}X_{n+1} = (n-1)^{\lambda a}X_n - \nu_n X_n - n^{\lambda a}R_n - n^{\lambda a}A_n E_n' .$$

Iterieren liefert:

$$(3) \qquad n^{\lambda a}X_{n+1} = -\sum_{i=1}^{n}(\nu_i X_i + i^{\lambda a}R_i) - \sum_{i=1}^{n} i^{\lambda a}A_i E_i' .$$

Nach Bemerkung iii) folgt aus Satz (3.2) und aus den Definitionen von R_n und ν_n:

$$\nu_n X_n + n^{\lambda a}R_n = O(n^{\lambda a-2-\delta} + n^{\lambda a-1-\delta-\eta\delta} + n^{\lambda a-1-\rho-\delta}) \text{ f.s.}$$

für alle $\delta < \Delta$. Insbesondere gibt es $\delta < \Delta$ mit $\delta(1+\eta) > \Delta$ und $\delta + \rho > \Delta$, so daß für $\varepsilon' := \min(\delta + \eta\delta - \Delta\ ,\ 1 + \delta - \Delta\ ,\ \rho + \delta - \Delta) > 0$ gilt:

$$n^{-\lambda a+\Delta}(\nu_n X_n + n^{\lambda a}R_n) = O(n^{-1-\varepsilon'}) \text{ f.s.}$$

Unter (E_1): $\lambda a = \Delta$ konvergiert somit $\Sigma_{n=1}^{\infty}(\nu_n X_n + n^{\lambda a} R_n)$ f.s., d.h. es existiert eine Zufallsvariavle Z mit

$$Z = -\Sigma_{n=1}^{\infty}(\nu_n X_n + n^{\lambda a} R_n) \text{ f.s.},$$

und aus (3) folgt die Behauptung (2).

Unter (E): $\lambda a > \Delta$ gibt es ε mit $0 < \varepsilon < \min(\lambda a - \Delta, \varepsilon')$. Für ein derartiges ε gilt erstens

$$\Sigma_{n=1}^{\infty} n^{-\lambda a+\Delta+\varepsilon}(\nu_n X_n + n^{\lambda a} R_n) \text{ konvergiert f.s.},$$

und zweitens können wir auf diese Reihe das dem Wahrscheinlichkeitstheoretiker wohlbekannte Lemma von Kronecker (vgl. Loève (1977) I, S. 25

anwenden und erhalten:

$$n^{-\lambda a+\Delta} \Sigma_{i=1}^{n}(\nu_i X_i + i^{\lambda a} R_i) = O(n^{-\varepsilon}) \text{ f.s.}$$

Nach Multipikation von (3) mit $n^{-\lambda a+\Delta}$ folgt daraus die Behauptung (1). □

Im nichtdeterministischen Falle wirken sich die zufälligen Gewichte A_n in der durch Satz (4.1) gegebenen starken Repräsentation noch störend aus. Jedoch erhalten wir unter einer zusätzlichen Forderung an das Verhalten von $\Sigma_{n=1}^{\infty} n^{\Delta} A_n E_n'$, die im deterministischen Fall $A_n = an^{-1}$ trivialerweise erfüllt ist:

(4.2) Korollar:

Ist zusätzlich zu den Voraussetzungen (A) bis (D) des Satzes (4.1)

(F) $\Sigma_{n=1}^{\infty} n^{\Delta}(A_n - an^{-1})E_n'$ konvergiert f.s.

erfüllt, dann gilt unter (E): $\lambda a > \Delta$:

(1) $$n^{\Delta}(X_{n+1} - \theta) = -an^{-\lambda a+\Delta} \Sigma_{i=1}^{n} i^{\lambda a-1} E_i' + o(1) \text{ f.s.},$$

und unter (E_1): $\lambda a = \Delta$ existiert eine Zufallsvariable Z' , so daß

(2) $$n^{\lambda a}(X_{n+1} - \theta) = - a \Sigma_{i=1}^{n} i^{\lambda a-1} E_i' + Z' + o(1) \text{ f.s.}$$

gilt.

Bemerkung:

Aus (D) und (F) folgt die fast sichere Konvergenz von $\Sigma_{n=1}^{\infty} n^{-1+\delta} E_n'$ für alle $\delta < \Delta$. Im allgemeinen ist also $\Delta > \frac{1}{2}$ nicht zu erwarten, sondern bestenfalls $\Delta = \frac{1}{2}$ (vgl. Bemerkung v) zu Satz (3.2)).

Beweis:

Die Behauptung (1) ergibt sich nach Anwendung des Lemmas von Kronecker auf (F):

$$n^{-\lambda a+\Delta} \Sigma_{i=1}^{n} i^{\lambda a}(A_i - ai^{-1})E_i' \to 0 \text{ f.s.},$$

während die Behauptung (2) sofort mit

$$Z' = Z - \Sigma_{n=1}^{\infty} n^{\lambda a}(A_n - an^{-1})E_n' \text{ f.s.}$$

aus Satz (4.1) folgt. □

Nachdem wir mit theoretischem Rüstzeug versehen sind, wollen wir uns "praktischen" Aufgaben zuwenden.

Kapitel 5: Klassische stochastische Approximation

In diesem Kapitel werden wir uns mit den in der Einleitung vorgestellten, speziellen stochastischen Approximationsprozessen mit deterministischer Schrittweite $A_n = an^{-1}$, $a > 0$, befassen. Zu jedem einzelnen Prozeß werden wir zuerst unter der recht schwachen Forderung

$$\Sigma_{n=1}^{\infty} n^{-1+\delta} U_n \quad \text{konvergiert f.s. für alle } \delta < \frac{1}{2}$$

eine starke Repräsentation durch gewichtete Fehlersummen formulieren, wobei die U_n zufällige Störungen in den Beobachtungen der Funktion M zum Zeitpunkt n sind. Wie bereits früher bemerkt, ist im allgemeinen Konvergenz der Summe für $\delta \geq \frac{1}{2}$ nicht zu erwarten.

Diese Forderung ist zum Beispiel dann erfüllt, wenn die Fehler U_n einem fast sicheren Invarianzprinzip folgender Form genügen:

Es existieren $\sigma > 0$ und eine Brown'sche Bewegung B, so daß gilt:

$$(5.1) \qquad \Sigma_{i \leq t} U_i = \sigma B(t) + o(t^{\frac{1}{2}}) \quad \text{f.s.}$$

(vgl. Anhang: Satz (A.4)). Eine andere Beispielsklasse werden wir im nächsten Kapitel kennenlernen.

Die Gültigkeit derartiger Invarianzprinzipien wurde erstmals von Strassen (1964) und in der hier vorliegenden verschärften Form mit einem Abweichungsterm der Größenordnung $o(t^{\frac{1}{2}})$ von Strassen (1966) für Summen unabhängiger Zufallsvariablen und für Martingale unter einer der Lindeberg-Bedingung verwandten Forderung bewiesen. Aber auch für etliche schwach abhängige Folgen $(U_n)_{n \in \mathbb{N}}$ ist (5.1) erfüllt (vgl. Philipp und Stout (1975) und Berkes und Philipp (1979)). Dabei ist

zu beachten, daß man notfalls die U_n , $n \in \mathbb{N}$ - und eventuell auftretende zusätzliche Zufallsvariablen - zusammen mit einer Brown'schen Bewegung B auf einem geeignet reichhaltigen Wahrscheinlichkeitsraum neu definieren muß, ohne ihre gemeinsame Verteilung zu ändern. Dies ist aber ohne weiteres möglich (vgl. Berkes und Philipp (1979), Lemma A.1). Weiterhin ist dieser Übergang zu einem anderen Wahrscheinlichkeitsraum ohne Auswirkungen auf unsere Aussagen, da uns lediglich der Bildraum der Zufallsvariablen interessiert.

Unter der Annahme eines fast sicheren Invarianzprinzips vom Typ (5.1) ist es uns möglich, zu konkreteren Ergebnissen wie asymptotische Normalität, Gesetze vom iterierten Logarithmus oder gar Invarianzprinzipien für die betrachteten stochastischen Approximationsprozesse zu gelangen. Die dazu benutzten Erkenntnisse über Folgen $(U_n)_{n\in\mathbb{N}}$, die einem fast sicheren Invarianzprinzip genügen, werden wir ausführlich im zweiten Teil des Anhangs darstellen.

Der Robbins-Monro-Prozeß

(5.2) Satz:

Genügen $M: \mathbb{R} \to \mathbb{R}$, $\vartheta \in \mathbb{R}$, Zufallsvariablen X_1 und U_n , $n \in \mathbb{N}$, den Voraussetzungen:

A) $M(x)(x - \vartheta) > 0$ für alle $x \neq \vartheta$,

B) $\inf_{\delta_1 \le |x-\vartheta| \le \delta_2} |M(x)| > 0$ für alle $\delta_2 \ge \delta_1 > 0$,

C) es ex. $\alpha, \beta > 0$ mit $|M(x)| \le \alpha + \beta|x - \vartheta|$ für alle x ,

D) $\sum_{n=1}^{\infty} n^{-1} U_n$ konvergiert f.s. ,

dann gilt für den durch

$$X_{n+1} = X_n - an^{-1}Y_n$$

mit

$$Y_n = M(X_n) + U_n \quad ,$$

$a > 0$, definierten Robbins-Monro-Prozeß:

(1) $\quad X_n \to \theta$ f.s.

Ist zusätzlich

(E) M ist differenzierbar in θ mit $M'(\theta) = \lambda > 0$, und es ex. $\eta > 0$ mit $M(x) = M'(\theta)(x - \theta) + O(|x - \theta|^{1+\eta})$ für $x \to \theta$

und an Stelle von (D) die stärkere Voraussetzung

(D') $\Sigma_{n=1}^{\infty} n^{-1+\delta} U_n$ konvergiert f.s. für alle $\delta < \frac{1}{2}$

erfüllt, dann gilt

i) für $\lambda a > \frac{1}{2}$:

(2) $\quad n^{\frac{1}{2}}(X_{n+1} - \theta) = - an^{-\lambda a + \frac{1}{2}} \Sigma_{i=1}^{n} i^{\lambda a - 1} U_i + O(n^{-\varepsilon})$ f.s.

für ein $\varepsilon > 0$;

ii) für $\lambda a = \frac{1}{2}$:

es existiert eine Zufallsvariable Z mit

(3) $\quad n^{\frac{1}{2}}(X_{n+1} - \theta) = - a\Sigma_{i=1}^{n} i^{-\frac{1}{2}} U_i + Z + o(1)$ f.s. ;

iii) für $\lambda a < \frac{1}{2}$:

es existiert eine Zufallsvariable Z* mit

(4) $\quad n^{\lambda a}(X_{n+1} - \theta) \to Z^*$ f.s.

Beweis:

Mit $f := M$, $A_n := an^{-1}$, $E_n := U_n$ und $B_n := 0$ besitzt der Robbins-Monro-Prozeß die Form (2.1), und die Behauptung (1) folgt, da die Voraussetzungen (A), (B), (C) und (E) bzw. (D), (F) und (G) des Satzes (2.2) auf Grund von (A) bis (D) bzw. auf Grund der Definition von A_n und B_n erfüllt sind.

Setzen wir weiterhin $E_n' := E_n + B_n = U_n$, dann sind für $\Delta = \min(\frac{1}{2}, \lambda a)$ die Voraussetzungen des Satzes (4.1) erfüllt, und die Behauptungen (2) und (3) folgen sofort. Behauptung (4) für $\lambda a < \frac{1}{2}$ ergibt sich unter nochmaliger Beachtung von (D') mit

$$Z^* = Z - a\sum_{n=1}^{\infty} n^{-1+\lambda a} U_n \quad \text{f.s.} \qquad \square$$

Im Falle $aM'(\theta) < \frac{1}{2}$ ist die Konvergenzrate deutlich schlechter als $n^{-\frac{1}{2}}$, und wir wollen uns daher mit dem Ergebnis (5.2) (4) begnügen, das im wesentlichen nur Satz 2 von Major und Révész (1973) übereinstimmt. In den Fällen $aM'(\theta) > \frac{1}{2}$ und $aM'(\theta) = \frac{1}{2}$ erhalten wir unter der Annahme eines fast sicheren Invarianzprinzips:

(5.3) Korollar:

Sind die Voraussetzungen (A) bis (C) und (E) des Satzes (5.2) und an Stelle von (D') die stärkere Forderung

(D*) es existiert ein $\varepsilon > 0$ und eine Brown'sche Bewegung B, so daß gilt:

$$\max_{i \le t} U_i = \sigma B(t) + o(t^{\frac{1}{2}}) \quad \text{f.s.}$$

erfüllt, dann gilt für den Robbins-Monro-Prozeß

i) im Falle $\lambda a > \frac{1}{2}$

a) ein starkes Invarianzprinzip:

Es existiert eine Brown'sche Bewegung B' , so daß gilt:

$$(1) \qquad [t]^{\frac{1}{2}}(X_{[t]} - \theta) = a\sigma(2\lambda a - 1)^{-\frac{1}{2}} t^{-\lambda a + \frac{1}{2}} B'(t^{2\lambda a - 1}) + o(1) \quad \text{f.s.}$$

b) ein Gesetz vom iterierten Logarithmus:

$$(2) \qquad \limsup_{n \to \infty} \frac{n^{\frac{1}{2}}|X_n - \theta|}{\sqrt{2 \log\log n}} = a\sigma(2\lambda a - 1)^{-\frac{1}{2}} \quad \text{f.s.} \quad ;$$

c) ein Verteilungsinvarianzprinzip:

$$(3) \qquad (t^{\lambda a} n^{\frac{1}{2}}(X_{[nt]} - \theta))_{t \in [0,1]} \xrightarrow{\mathcal{D}} Z \quad (\text{in } \mathcal{D}[0,1]) \quad ,$$

wobei Z eine zufällige Funktion mit

$$Z(t) = a\sigma(2\lambda a - 1)^{-\frac{1}{2}} B(t^{2\lambda a - 1})$$

für $0 \leq t \leq 1$ und Brown'scher Bewegung B ist;

d) asymptotische Normalität:

$$(4) \qquad n^{\frac{1}{2}}(X_n - \theta) \xrightarrow{\mathcal{D}} Y$$

mit $Y \sim N(0, a^2\sigma^2(2\lambda a - 1)^{-1})$.

ii) im Falle $\lambda a = \frac{1}{2}$, wenn der Fehlerterm in (D*) höchstens von der Größenordnung $o(t^{\frac{1}{2}}(\log t)^{-\frac{1}{2}})$ f.s. ist:

a) ein starkes Invarianzprinzip:

Es existiert eine Brown'sche Bewegung B' ,

so daß gilt:

$$(5) \qquad [t]^{\frac{1}{2}}(X_{[t]} - \theta) = a\sigma B'(\log t) + o((\log t)^{\frac{1}{2}}) \quad \text{f.s.} \quad ;$$

b) ein Gesetz vom iterierten Logarithmus:

$$(6) \qquad \limsup_{n \to \infty} \frac{n^{\frac{1}{2}}|X_n - \theta|}{\sqrt{2 \log n \log\log\log n}} = a\sigma \quad \text{f.s.} \quad ;$$

c) asymptotische Normalität:

(7) $\quad n^{\frac{1}{2}}(\log n)^{-\frac{1}{2}}(X_n - \theta) \xrightarrow{\mathcal{D}} Y$

mit $Y \sim N(0, a^2\sigma^2)$.

Bemerkungen:

i) Das Auftreten von X_0 in (1) bzw. (5) und insbesondere in (3) ist rein ästhetischer Natur. Dabei kann X_0 beliebig gesetzt werden, z.B. $X_0 = 0$ oder $X_0 = X_1$.

ii) Sowohl aus den Gesetzen vom iterierten Logarithmus (2) bzw. (6) als auch aus den Aussagen über asymptotische Normalität (4) bzw. (7) ist ersichtlich, daß die Konvergenzrate im Falle $\lambda a = \frac{1}{2}$ schlechter als für $\lambda a > \frac{1}{2}$ ist. Im letzteren Fall wird insbesondere minimale Varianz der Grenzverteilung in (4) erreicht, falls $a = \lambda^{-1}$ gilt. Bei dieser speziellen Wahl von a ist $\mathrm{Var}(Y) = \sigma^2\lambda^{-2}$ und das im Verteilungsinvarianzprinzip auftretende Z ist - bis auf eine multiplikative Konstante - Brown'sche Bewegung. Im allgemeinen ist Z für $\lambda a > \frac{1}{2}$ ein zentrierter Gaußprozeß auf [0,1] mit Kovarianzfunktion $K(s,t) = a^2\sigma^2(2\lambda a - 1)^{-1}\,(t^{2\lambda a-1} - s^{2\lambda a-1})$, $s \le t$, also insbesondere mit unabhängigen Zuwächsen. Für den Fall $\lambda a = \frac{1}{2}$ können wir kein Verteilungsinvarianzprinzip formulieren.

iii) Setzen wir voraus, daß die Fehler U_n bedingt unabhängig sind (vgl. Kapitel 1), dann bilden sie eine Martingaldifferenzfolge. Sind weiterhin für ein $p > 1$ die p-ten absoluten Momente der F_x gleichmäßig beschränkt, d.h. $\int |y|^p F_x(dy) \le c_p < \infty$ für alle x , dann ist (5.2) (D) erfüllt, und im Falle gleichmäßig beschränkter Varianzen, d.h. $p = 2$

gilt sogar (5.2) (D') (vgl. dazu etwa Hall und Heyde (1980), S. 35). Unter den Voraussetzungen von Kersting (1977) und mit den von ihm beschriebenen Methoden läßt sich sogar ein fast sicheres Invarianzprinzip beweisen. Existiert zusätzlich das p-te absolute Moment von F_θ für ein $p > 2$, dann ist dieses Invarianzprinzip sogar von der Form (5.1) (zu diesem Thema vgl. Csörgö und Révész (1981), S. 94f, S. 107f). Diese Überlegungen mögen die Wahl der Voraussetzung (D*) rechtfertigen. Es sei noch bemerkt, daß für bedingt unabhängige Fehler das vorliegende Korollar eine Sammlung bekannter Ergebnisse darstellt (für Quellenangaben vgl. Kapitel 1).

<u>Beweis:</u>

Zu (1):

Auf [illegible] von (D*) gibt es nach Satz (A. 4) zu $\alpha := \lambda a - 1 > -\frac{1}{2}$ eine Brown'sche Bewegung B_α mit

$$\Sigma_{i \leq t} i^{\lambda a - 1} U_i = \sigma(2\lambda a - 1)^{-\frac{1}{2}} B_\alpha(t^{2\lambda a - 1}) + o(t^{\lambda a - \frac{1}{2}}) \quad \text{f.s.}$$

Wegen [illegible] $+ 1]^{\frac{1}{2}} = ([t] + 1)^{\frac{1}{2}} = [t]^{\frac{1}{2}} + O($[illegible] dank Mittelwertsatz und nach (5.2) (1) gilt weiterhin:

$$[t + 1]^{\frac{1}{2}}(X_{[t+1]} - \theta) = [t]^{\frac{1}{2}}(X_{[t]+1} - \theta) + o(1) \quad \text{f.s.}$$

Bea[illegible], daß mit B_α auch B' := [illegible]'sche Bewegung ist, so [illegible] (5.2) (2) mit Hilfe [illegible]gangener Überlegungen:

$$[t + 1]^{\frac{1}{2}}(X_{[t+1]} - \theta) = a\sigma(2\lambda a - 1)^{-\frac{1}{2}} [t]^{-\lambda a + \frac{1}{2}} B'(t^{2\lambda a - 1}) + \ldots$$

Dieses Ergebnis ist schon fast zufriedenstellend - wir wollen aber noch mehr: Durch geschicktes Addieren von null erhalten wir:

$$[t]^{-\lambda a+\frac{1}{2}}B'(t^{2\lambda a-1}) = (t+1)^{-\lambda a+\frac{1}{2}}B'((t+1)^{2\lambda a-1}) - R_1(t) - R_2(t)$$

mit

$$R_1(t) = (t+1)^{-\lambda a+\frac{1}{2}}\,(B'((t+1)^{2\lambda a-1}) - B'(t^{2\lambda a-1}))\quad ,$$

$$R_2(t) = ((t+1)^{-\lambda a+\frac{1}{2}} - [t]^{-\lambda a+\frac{1}{2}})\,B'(t^{2\lambda a-1})\quad .$$

Wegen $(t+1)^{2\lambda a-1} = t^{2\lambda a-1} + O(t^{2\lambda a-2})$ ergibt sich mit Hilfe des auch im Beweis von Satz (A.4) (vgl. S. 121) benutzten tieferliegenden Ergebnisses für die Zuwächse einer Brown'schen Bewegung (A.6)

$$B'((t+1)^{2\lambda a-1}) = B'(t^{2\lambda a-1}) + o(t^{\lambda a-\frac{1}{2}})\quad \text{f.s.}$$

und folglich:

$$R_1(t) = o(1)\quad \text{f.s.}$$

Weiterhin gilt $(t+1)^{-\lambda a+\frac{1}{2}} - [t]^{-\lambda a+\frac{1}{2}} = O(t^{-\lambda a-\frac{1}{2}})$, und mit dem Gesetz vom iterierten Logarithmus folgt

$$R_2(t) = o(1)\quad \text{f.s.}\quad .$$

Dieses in (8) eingesetzt liefert

$$[t+1]^{\frac{1}{2}}(X_{[t+1]} - \theta) = a\sigma(2\lambda a-1)^{-\frac{1}{2}}\,(t+1)^{-\lambda a+\frac{1}{2}}B'((t+1)^{2\lambda a-1}) + o(1)\quad \text{f.s.}\quad ,$$

woraus nach Übergang von $t+1$ zu 1 die Behauptung (1) folgt.

Im Beweis von (5) erhält man aus (5.2) (3) analog zu (8)

$$[t+1]^{\frac{1}{2}}(X_{[t+1]} - \theta) = a\sigma B'(\log t) + o((\log t)^{\frac{1}{2}})\quad \text{f.s.}\quad .$$

Da $\log(t+1) = \log t + O(t^{-1})$ gilt, folgt unter erneuter Verwendung von (A.6) (vgl. auch S. 121) die Behauptung (5).

(2) bzw. (6) erhalten wir nun sofort aus (1) bzw. (5) nach geeigneter Variablentransformation mit dem Gesetz vom iterierten Logarithmus.

Setzen wir $X(t) := (a\sigma)^{-1}(2\lambda a - 1)^{\frac{1}{2}} t^{\lambda a}(X_{[t]} - \theta)$, dann folgt aus (1) unter Beachtung von $t^{\lambda a} = t^{\lambda a - \frac{1}{2}}[t]^{\frac{1}{2}} + O(t^{\lambda a - 1})$ und (5.2) (1):

$$X(t) = B'(t^{2\lambda a - 1}) + o(t^{\lambda a - \frac{1}{2}}) \quad \text{f.s.}$$

Mit Satz (A. 5) erhalten wir

$$(n^{-\lambda a + \frac{1}{2}} X(nt))_{t \in [0,1]} \xrightarrow{\mathcal{D}} (B(t^{2\lambda a - 1}))_{t \in [0,1]} \quad (\text{in } \mathcal{D}[0,1]) \ ,$$

was nach Einsetzen von X(nt) und Multiplikation mit $a\sigma(2\lambda a - 1)^{\frac{1}{2}}$ gerade liefert.

(4) ergibt sich sofort durch Auswertung von (3) an der Stelle t = mit Y = Z(1) (vgl. Billingsley (1968), S. 30, S. 121). Für (7) müssen wir lediglich beachten, daß $a\sigma(\log n)^{-\frac{1}{2}} B'(\log n) \sim N(0, a^2\sigma^2)$, $n \in \mathbb{N}$, und nach (5)

$$n^{\frac{1}{2}}(\log n)^{-\frac{1}{2}}(X_n - \theta) - a\sigma(\log n)^{-\frac{1}{2}} B'(\log n) \to 0 \quad \text{f.s.}$$

gilt, und aus diesen beiden Aussagen folgt die Behauptung (vgl. Billi (1968), S. 25). □

Der Kiefer-Wolfowitz-Prozeß

(5.4) Satz:

Genügen $M: \mathbb{R} \to \mathbb{R}$, $\theta \in \mathbb{R}$, Zufallsvariablen X_1 und U_n , $n \in \mathbb{N}$, den Voraussetzungen:

(A) M ist stetig differenzierbar und $M'(x)(x - \theta) > 0$ für alle $x \neq \theta$,

(B) es ex. $\alpha,\beta > 0$ mit $|M'(x)| \leq \alpha + \beta|x - \theta|$ für alle x ,

(C) $\Sigma_{n=1}^{\infty} n^{-1+\delta} U_n$ konvergiert f.s. für alle $\delta < \frac{1}{2}$,

dann gilt für den durch

$$X_{n+1} = X_n - an^{-1}Y_n$$

mit

$$Y_n = (2c_n)^{-1}(M(X_n + c_n) - M(X_n - c_n) + U_n) \quad ,$$

$a > 0$, $c_n = cn^{-\gamma}$, $c > 0$, $0 < \gamma < \frac{1}{2}$, definierten Kiefer-Wolfowitz-Prozeß:

(1) $X_n \to \theta$ f.s.

Ist zusätzlich

(D) M ist dreimal differenzierbar in einer Umgebung U von θ ,
und M''' ist stetig in θ ,

(E) $aM'(\theta) > \frac{1}{2} - \gamma$

erfüllt, dann gilt

i) für $\gamma = \frac{1}{6}$:

(2) $$n^{\frac{1}{3}}(X_{n+1} - \theta) = \mu - a(2c)^{-1}n^{-aM'(\theta)+\frac{1}{3}} \Sigma_{i=1}^{n} i^{aM'(\theta)-\frac{5}{6}} U_i + o(1) \quad \text{f.s.}$$

mit $\mu := -(6aM'(\theta) - 2)^{-1}ac^2M'''(\theta)$;

ii) für $\gamma > \frac{1}{6}$:

(3) $$n^{\frac{1}{2}-\gamma}(X_{n+1} - \theta) = -a(2c)^{-1}n^{-aM'(\theta)+\frac{1}{2}-\gamma} \Sigma_{i=1}^{n} i^{aM'(\theta)-1+\gamma} U_i + o(1) \quad \text{f.s.} \quad ;$$

iii) für $\gamma < \frac{1}{6}$:

$$(4) \qquad n^{2\gamma}(X_{n+1} - \theta) = -\,(6aM'(\theta) - 12\gamma)^{-1}ac^2M'''(\theta) + o(1) \quad \text{f.s.}$$

Bemerkungen:

i) Wie wir den Ergebnissen (2) bis (4) entnehmen können, ist die Wahl von $\gamma = \frac{1}{6}$ in dem Sinne optimal, daß für $\gamma > \frac{1}{6}$ schon beim Vorliegen von Fehlern, die einem fast sicheren Invarianzprinzip (5.1) genügen, und für $\gamma < \frac{1}{6}$ im Falle $M'''(\theta) \neq 0$ schlechtere Konvergenzraten auftreten. Ähnlich wie beim Robbins-Monro-Prozeß verschlechtert sich die Konvergenzgeschwindigkeit auch dann, wenn (E) verletzt ist.

ii) Statt der Differenzierbarkeit von M genügt es, diese lediglich stückweise und die Existenz linksseitiger und rechtsseitiger Ableitunge zu fordern sowie (A) und (B) durch entsprechende Voraussetzungen an die einseitigen Ableitungen zu ersetzen. Damit können dann auch Funktionen mit "Knicken" wie M mit $M(x) = |x|$ betrachtet werden.

Beweis:

Mit $A_n := an^{-1}$, $E_n := (2c_n)^{-1}U_n$ und f_n definiert durch $f_n(x) := (2c_n)^{-1}(M(x + c_n) - M(x - c_n))$ besitzt der Kiefer-Wolfowitz-Prozeß die Form (2.3), und die Voraussetzungen (D) und (E) von Satz (2.4) sind nach Definition von A_n und auf Grund von $A_nE_n = a(2c)^{-1}n^{-1+\gamma}$ nach (C) offensichtlich erfüllt. Dank Mittelwertsatz existiert für alle x und alle n ein $\xi(x,n)$ mit $|\xi(x,n) - x| < c_n$ und

$$f_n(x) = M'(\xi(x,n)) \quad .$$

Weiterhin gibt es zu $\varepsilon > 0$ ein $N(\varepsilon)$, so daß $c_n < \varepsilon$ für $n \geq N(\varepsilon)$ gilt. Dann besitzen aber x mit $|x - \theta| > \varepsilon$ und $\xi(x,n)$ für $n \geq N(\varepsilon)$ dasselbe Vorzeichen, so daß (2.4) (A) über $M'(\xi(x,n))(x - \theta) > 0$ aus (A) folgt. Ist $\varepsilon < \delta_1$, dann gilt für $n \geq N(\varepsilon)$

$$\inf_{\delta_1 \leq |x-\theta| \leq \delta_2} |f_n(x)| \geq \inf_{\delta_1 - \varepsilon \leq |\xi-\theta| \leq \delta_2 + \varepsilon} |M'(\xi)| \quad ,$$

und die rechte Seite ist auf Grund der Stetigkeit von M' und wegen $M'(x)(x - \theta) > 0$ positiv, so daß auch (2.4) (B) erfüllt ist. (2.4) (C) ergibt sich über

$$|f_n(x)| = |M'(\xi(x,n))| \leq \alpha + \beta|\xi(x,n) - \theta| \leq \alpha + \beta c_n + \beta|x - \theta|$$

aus (B) mit $\alpha + \beta c$ an Stelle von α . Satz (2.4) ist somit anwendbar und liefert die fast sichere Konvergenz (1).

Wir setzen nun $f := M'$ und $E_n' := E_n + B_n$ mit $B_n := f_n(X_n) - f(X_n)$. Wegen $E_n = (2c)^{-1} n^{\gamma} U_n$ und $A_n = an^{-1}$ folgt aus (C)

(5) $\quad \Sigma_{n=1}^{\infty} n^{\delta} A_n E_n$ konvergiert f.s. für alle $\delta < \frac{1}{2} - \gamma$.

Als nächstes wollen wir zeigen:

(6) $\quad n^{2\gamma} B_n \to \frac{1}{6} c^2 M'''(\theta)$ f.s.

Dazu betrachten wir wieder Realisationen:

Sei also $\omega \in \Omega' := \{\omega | \lim_{n\to\infty} X_n(\omega) = \theta\}$. Nach (1) ist $P(\Omega') = 1$. Wegen $\lim_{n\to\infty} x_n = \theta$ und $\lim_{n\to\infty} c_n = 0$ existiert ein N , so daß für alle $n \geq N$ das Intervall $[x_n - c_n , x_n + c_n]$ in der Umgebung U von θ liegt. Nach (D) ist somit M für $n \geq N$ auf $[x_n - c_n , x_n + c_n]$ dreimal differenzierbar, und wir können nach Taylor um x_n entwickeln:

$$M(x_n \pm c_n) = M(x_n) \pm c_n M'(x_n) + \frac{1}{2} c_n^2 M''(x_n) \pm \frac{1}{6} c_n^3 M'''(\xi_n^{\pm}) \quad ,$$

für gewisse ξ_n^-, ξ_n^+ mit $x_n - c_n < \xi_n^- < x_n < \xi_n^+ < x_n + c_n$ (wobei diese Aussage wie auch ähnliche später auftretende für "+" und "-" getrennt zu lesen ist). Daraus folgt

$$M(x_n + c_n) - M(x_n - c_n) = 2c_n M'(x_n) + \frac{1}{6} c_n^3 (M'''(\xi_n^-) + M'''(\xi_n^+))$$

und damit:

$$b_n = (2c_n)^{-1}(M(x_n + c_n) - M(x_n - c_n)) - M'(x_n)$$
$$= \frac{1}{12} c_n^2 (M'''(\xi_n^-) + M'''(\xi_n^+)) \quad .$$

Nun gilt aber $\lim_{n\to\infty} \xi_n^- = \lim_{n\to\infty} \xi_n^+ = \theta$, und aus der Stetigkeit von M''' in θ folgt unter Beachtung von $c_n^2 = c^2 n^{-2\gamma}$ die Behauptung.

Aus (6) ergibt sich sofort

(7) $\quad \Sigma_{n=1}^{\infty} n^{\delta} A_n B_n$ konvergiert f.s. für alle $\delta < 2\gamma$.

Wegen (5) und (7) ist somit Voraussetzung (4.1) (D) für $\Delta = \min(2\gamma, \frac{1}{2} -$ erfüllt. Nach (E) gilt $\lambda = f'(\theta) = M''(\theta) > 0$ und $\lambda a > \Delta$. Da die Stetigkeit von M''' in θ die Beschränktheit von M''' in einer Umgebung $U' \subseteq U$ von θ nach sich zieht, ist nach Bemerkung ii) zu Satz (4.1) auch Voraussetzung (4.1) (A) erfüllt, so daß wir diesen Satz anwenden dürfen:

(8) $\quad n^{\Delta}(X_{n+1} - \theta) = R_1 + R_2 + o(1)$ f.s.

mit

$$R_1 = - a n^{-aM''(\theta)+\Delta} \Sigma_{i=1}^{n} i^{aM''(\theta)-1} B_i \quad ,$$

$$R_2 = - a(2c)^{-1} n^{-aM''(\theta)+\Delta} \Sigma_{i=1}^{n} i^{aM''(\theta)-1+\gamma} U_i \quad .$$

Für $\gamma > \frac{1}{6}$ gilt $\Delta = \frac{1}{2} - \gamma < 2\gamma$, und aus der fast sicheren Konvergenz von $\Sigma_{n=1}^{\infty} n^{\Delta-1} B_n$ (vgl. (7)) folgt wegen $aM''(\theta) - \Delta > 0$ mit dem Lemma von Kronecker

$$R_1 \to 0 \quad \text{f.s.}$$

und damit Behauptung (3).

Für $\gamma \leq \frac{1}{6}$ gilt $\Delta = 2\gamma$. Zur Untersuchung von R_1 wollen wir das Lemma von Toeplitz (vgl. Loève (1977) I, S. 250) anwenden, welches häufig zum Beweis des Lemmas von Kronecker benutzt wird:

Nach Lemma (A. 1) gilt wegen $aM''(\theta) > \Delta$:

$$\lim_{n\to\infty} \Sigma_{i=1}^{n} n^{-aM''(\theta)+\Delta_j aM''(\theta)-\Delta-1} = (aM''(\theta) - \Delta)^{-1} \quad .$$

Da alle Summanden positiv sind, folgt nun aus (6) mit dem Lemma von Toeplitz

$$R_1 \to - a(aM''(\theta) - \Delta)^{-1} \frac{1}{6} c^2 M'''(\theta) \quad \text{f.s.}$$

Für $\gamma = \frac{1}{6}$ ergibt sich nun Behauptung (2) durch Einsetzen in (8).

Für $\gamma < \frac{1}{6}$ folgt aus der f.s. Konvergenz von $\Sigma_{n=1}^{\infty} n^{-1+3\gamma} U_n$ (vgl. (C)) wegen $aM''(\theta) > \Delta$ mit dem Lemma von Kronecker:

$$R_2 \to 0 \quad \text{f.s.} \quad ,$$

und die Behauptung (4) ergibt sich wieder durch Einsetzen in (8). □

Unter der Annahme eines fast sicheren Invarianzprinzips erhalten wir für den "optimalen" Fall $\gamma = \frac{1}{6}$:

(5.5) Korollar:

Sind die Voraussetzungen des Satzes (5.4) erfüllt, und statt

(C) gelte die stärkere Forderung

(C*) es ex. $\sigma > 0$ und eine Brown'sche Bewegung B , so daß gilt:

$$\Sigma_{i\leq t} U_i = \sqrt{2}\, \sigma B(t) + o(t^{\frac{1}{2}}) \quad \text{f.s.} \quad ,$$

dann gilt für den Kiefer-Wolfowitz-Prozeß mit $\gamma = \frac{1}{6}$

a) ein starkes Invarianzprinzip:

Es existiert eine Brown'sche Bewegung B' , so daß gilt:

$$(1) \qquad [t]^{\frac{1}{3}}(X_{[t]} - \theta) = \mu + a\sigma(2c)^{-1}(aM''(\theta) - \tfrac{1}{3})^{-\frac{1}{2}} t^{-aM''(\theta)+\frac{1}{3}} B'(t^{2aM''(\theta)} + o(1) \quad \text{f.s.} \quad ;$$

b) ein Gesetz vom iterierten Logarithmus:

$$(2) \qquad \limsup_{n\to\infty} \frac{n^{\frac{1}{3}}|X_n - \theta|}{\sqrt{2\ \text{loglog}n}} = a\sigma(2c)^{-1}(aM''(\theta) - \tfrac{1}{3})^{-\frac{1}{2}} \quad \text{f.s.} \quad ;$$

c) ein Verteilungsinvarianzprinzip:

$$(3) \qquad (t^{aM''(\theta)} n^{\frac{1}{3}}(X_{[nt]} - \theta))_{t\in[0,1]} \xrightarrow{\mathcal{D}} Z \quad (\text{in } \mathcal{D}[0,1]) \quad ,$$

wobei Z eine zufällige Funktion mit

$$Z(t) = \mu t^{aM''(\theta) - \frac{1}{3}} + a\sigma(2c)^{-1}(aM''(\theta) - \tfrac{1}{3})^{-\frac{1}{2}} B(t^{2aM''(\theta) - \frac{2}{3}})$$

für $0 \le t \le 1$ und Brown'scher Bewegung B ist;

d) asymptotische Normalität:

$$(4) \qquad n^{\frac{1}{3}}(X_n - \theta) \xrightarrow{\mathcal{D}} Y$$

mit $Y \sim N(\mu, a^2\sigma^2(2c)^{-2}(aM''(\theta) - \frac{1}{3})^{-1})$.

(Dabei ist $\mu := -(6aM''(\theta) - 2)^{-1} ac^2 M'''(\theta)$ wie in Satz (5.4) definiert.)

Bemerkung:

Der Faktor $\sqrt{2}$ in (C*) wird lediglich eingeführt, um Übereinstimmung mit dem Fall der bedingten Unabhängigkeit zu erzielen. Haben etwa die F_x konstante Varianz σ^2 , dann besitzt U_n , das sich aus den bedingt unabhängigen Fehlern U_n^+ und U_n^- zusammensetzt: $U_n = U_n^+ - U_n^-$ (vgl. Kapitel 1), die Varianz $2\sigma^2$.

Beweis:

Auf Grund von (C*) gibt es nach Satz (A.4) zu $\alpha := aM''(\theta) - \frac{5}{6} > -\frac{1}{2}$ eine Brown'sche Bewegung B_α mit

$$\Sigma_{i\le t} i^{aM''(\theta)-\frac{5}{6}} U_i = 2^{-\frac{1}{2}}\sigma(2aM''(\theta) - \tfrac{2}{3})^{-\frac{1}{2}} B_\alpha(t^{2aM''(\theta)-\frac{2}{3}})$$

$$+ o(t^{aM''(\theta)-\frac{1}{3}}) \quad \text{f.s.}$$

Aus (5.4) (2) folgt analog zu (5.3) (8):

$$[t+1]^{\frac{1}{3}}(X_{[t+1]} - \theta)$$

$$= \mu + a\sigma(2c)^{-1}(aM''(\theta) - \tfrac{1}{3})^{-\frac{1}{2}}[t]^{-aM''(\theta)+\frac{1}{3}} B'(t^{2aM''(\theta)-\frac{2}{3}}) + o(1) \quad \text{f.s.}$$

mit $B' = -B_\alpha$, und mit denselben Überlegungen wie im Beweis von (5.3) (1) ergibt sich die Behauptung (1).

Beachtet man noch für (2), daß $\lim_{n\to\infty} \mu(\log\log n)^{-\frac{1}{2}} = 0$ gilt, und für (3), daß aus $Y_n \xrightarrow{\mathcal{D}} Y$ (in $\mathcal{D}[0,1]$) für $h \in \mathcal{D}[0,1]$ auch $Y_n + h \xrightarrow{\mathcal{D}} Y + h$ (in $\mathcal{D}[0,1]$) folgt (vgl. Billingsley (1968), S. 30), dann ergeben sich die restlichen Aussagen wie in (5.3). □

Wie bereits bemerkt, ist die Wahl von $\gamma = \frac{1}{6}$ dann optimal, wenn $M'''(\theta) \neq 0$ gilt. Im Falle $M'''(\theta) = 0$ können wir jedoch unter zusätzlichen Voraussetzungen an höhere Ableitungen von M für andere Werte von γ bessere Konvergenzraten erhalten.

(5.6) Satz:

Sind die Voraussetzungen (A) bis (C) und (E) des Satzes (5.4) erfüllt und gilt für eine natürliche Zahl k:

(D*) M ist $(2k+1)$-mal differenzierbar in einer Umgebung U von θ,

$M^{(2k+1)}$ ist stetig in θ und $M^{(2j+1)}(\theta) = 0$ für $j = 1,\dots,k-1$,

dann gilt für den Kiefer-Wolfowitz-Prozeß für $\gamma = \gamma(k) := (4k + 2)^{-1}$:

$$(1) \qquad n^{\frac{1}{2}-\gamma}(X_{n+1} - \theta) = \mu(k,\gamma) + S(\gamma) + o(1) \quad \text{f.s.} \ ,$$

für $\gamma > \gamma(k)$:

$$(2) \qquad n^{\frac{1}{2}-\gamma}(X_{n+1} - \theta) = S(\gamma) + o(1) \quad \text{f.s.}$$

und für $\gamma < \gamma(k)$:

$$(3) \qquad n^{2k\gamma}(X_{n+1} - \theta) = \mu(k,\gamma) + o(1) \quad \text{f.s.}$$

mit

$$\mu(k,\gamma) := - ((2k + 1)!(aM'(\theta) - 2k\gamma))^{-1} a c^{2k} M^{(2k+1)}(\theta) \ ,$$

$0 < \gamma \leq \gamma(k)$, und

$$S(\gamma) := - a(2c)^{-1} n^{-aM'(\theta)+\frac{1}{2}-\gamma} \sum_{i=1}^{n} i^{aM'(\theta)-1+\gamma} U_i \ .$$

Bemerkungen:

i) Da $\frac{1}{2} - \gamma(k) = 2k\gamma(k)$ für $\gamma(k) = (4k + 2)^{-1}$ gilt, ist diese Wahl von $\gamma(k)$ wieder in dem Sinne optimal, daß sich für größere γ sofort und für kleinere im Falle $M^{(2k+1)}(\theta) \neq 0$ eine schlechtere Konvergenzrate ergibt. Für $k = 1$ entsprechen die Aussagen des vorliegenden Satzes gerade denen des Satzes (5.4), und mit wachsendem k wird $\gamma(k)$ kleiner, also die Konvergenzgeschwindigkeit größer.

ii) Unter der Voraussetzung eines starken Invarianzprinzips der Form (5.5) (C*) für die Fehler U_n können wir natürlich auch unter (D*) dem Korollar (5.5) entsprechende Ergebnisse beweisen, was wir hier aber nicht weiter ausführen werden. Als Beispiel sei nur erwähnt, daß für den Kiefer-Wolfowitz-Prozeß mit $\gamma = \gamma(k)$ unter (D*) $n^{\frac{1}{2}-\gamma(k)}(X_n -$ asymptotisch normalverteilt mit Erwartungswert $\mu(k,\gamma(k))$ und Varianz $a^2\sigma^2(2c)^{-2}(aM'(\theta) - \frac{1}{2} + \gamma(k))^{-1}$ ist.

Beweis:

Wir werden einen Induktionsbeweis durchführen:

Für $k = 1$ ist die Aussage des Satzes bereits bewiesen (vgl. Satz (5.4)). Unter der Induktionsvoraussetzung, daß die Aussagen des Satzes für $k - 1 \geq 1$ gelten, wollen wir zeigen, daß er auch für k richtig ist. Die Voraussetzung (D*) für k impliziert (D*) formuliert für $k - 1$ mit der Zusatzinformation $M^{(2k-1)}(\theta) = 0$. Nach Induktionsvoraussetzung gelten also die Aussagen (1) bis (3) formuliert für $k - 1$. Beachten wir noch, daß $\mu(k - 1, \gamma) = 0$ wegen $M^{(2k-1)}(\theta) = 0$ gilt, dann erhalten wir für $\gamma \geq \gamma(k - 1)$:

$$(4) \qquad n^{\frac{1}{2}-\gamma}(X_{n+1} - \theta) = S(\gamma) + o(1) \quad \text{f.s.}$$

und für $\gamma < \gamma(k - 1)$:

$$(5) \qquad X_n - \theta = o(n^{-2(k-1)\gamma}) \quad \text{f.s.}$$

Für $\gamma \geq \gamma(k - 1) > \gamma(k)$ haben wir also bereits das gewünschte Ergebnis, und für $\gamma < \gamma(k - 1)$ werden wir mit der Bezeichnungsweise von Satz (5.4) zeigen:

$$(6) \qquad n^{2k\gamma}B_n \to ((2k + 1)!)^{-1}c^{2k}M^{(2k+1)}(\theta) \quad \text{f.s.}$$

Damit erhalten wir wie im Beweis von (5.4) zuerst eine starke Repräsentation für $n^{\Delta}(X_{n+1} - \theta)$ der Gestalt (5.4) (8) mit $\Delta = \min(\frac{1}{2} - \gamma, 2k\gamma)$ und daraus mit der gleichen Argumentation über das Lemma von Kronecker bzw. das Lemma von Toeplitz die Aussagen (1) und (3) sowie (2) für den verbliebenen Fall $\gamma(k) < \gamma < \gamma(k - 1)$. Es bleibt noch, (6) zu beweisen: Der Weg hierzu geht über Realisationen $\omega \in \Omega' := \{\omega \mid \lim_{n\to\infty} n^{2(k-1)\gamma}(X_n(\omega) - \theta) = 0\}$, und auf Grund von (5) gilt $P(\Omega') = 1$. Für ein derartiges ω gibt es nach (D*) ein N, so daß $x_n - c_n$ und $x_n + c_n$ für $n \geq N$ in einem Intervall

um θ liegen, auf dem M $(2k+1)$-mal differenzierbar ist. Offensichtlich liegt auch x_n in diesem Intervall, und wir können M um x_n nach Taylor entwickeln:

$$M(x_n \pm c_n) = M(x_n) + \Sigma_{\ell=1}^{2k} \frac{1}{\ell!}(\pm c_n)^{\ell} M^{(\ell)}(x_n) \pm \frac{1}{(2k+1)!} c_n^{2k+1} M^{(}$$

für gewisse ξ_n^+, ξ_n^- mit $x_n - c_n < \xi_n^- < x_n < \xi_n^+ < x_n + c_n$. Damit ergibt si

$$M(x_n + c_n) - M(x_n - c_n) = 2c_n M'(x_n) + \Sigma_{j=1}^{k-1} \frac{2}{(2j+1)!} c_n^{2j+1} M^{(2j}$$

$$+ \frac{1}{(2k+1)!} c_n^{2k+1} (M^{(2k+1)}(\xi_n^+) + M^{(2k}$$

und folglich:

$$(7) \qquad b_n = (2c_n)^{-1}(M(x_n + c_n) - M(x_n - c_n)) - M'(x_n)$$

$$= \Sigma_{j=1}^{k-1} \frac{1}{(2j+1)!} c_n^{2j} M^{(2j+1)}(x_n)$$

$$+ \frac{1}{2(2k+1)!} c_n^{2k} (M^{(2k+1)}(\xi_n^+) + M^{(2k+1)}(\xi_n^-)) \quad .$$

Wir können nun für $1 \le j < k$ nach (D*) den Mittelwertsatz auf $M^{(2j+1)}$ anwenden und erhalten wegen $\lim_{n\to\infty} x_n = \theta$:

$$M^{(2j+1)}(x_n) = M^{(2j+1)}(\theta) + O'x_n - \theta) \quad .$$

Unter Berücksichtigung von $M^{(2j+1)}(\theta) = 0$ ergibt sich mit (5):

$$c_n^{2j} M^{(2j+1)}(x_n) = o(n^{-2j\gamma-2(k-1)\gamma}) \quad ,$$

so daß wegen $j \ge 1$ die in (7) auftretende Summe von der Größenordnung $o(n^{-2k\gamma})$ ist. Beachten wir noch $c_n^{2k} = c^{2k} n^{-2k\gamma}$ und $\lim_{n\to\infty} \xi_n^+ = \lim_{n\to\infty} \xi_n^- = \theta$ dann erhalten wir unter Ausnutzung der Stetigkeit von $M^{(2k+1)}$ in θ aus (7) die Behauptung (6). □

Burkholder's A_3-Prozeß

(5.7) Satz:

Genügen $M: \mathbb{R} \to \mathbb{R}$, $\theta \in \mathbb{R}$ und Zufallsvariablen X_1 und U_n den Voraussetzungen:

(A) M ist zweimal stetig differenzierbar und $M''(x)(x - \theta) > 0$ für alle $x \neq \theta$,

(B) es ex. $\alpha, \beta > 0$ mit $|M''(x)| \leq \alpha + \beta|x - \theta|$ für alle x ,

(C) $\sum_{n=1}^{\infty} n^{-1+\delta} U_n$ konvergiert f.s. für alle $\delta < \frac{1}{2}$

erfüllt, dann gilt für den durch

$$X_{n+1} = X_n - an^{-1}Y_n$$

mit

$$Y_n = c_n^{-2}(M(X_n + c_n) + M(X_n - c_n) - 2M(X_n) + U_n) \quad ,$$

$a > 0$, $c_n = cn^{-\gamma}$, $c > 0$, $0 < \gamma < \frac{1}{4}$, definierten A_3-Prozeß:

(1) $X_n \to \theta$ f.s.

Ist zusätzlich

(D) M ist viermal differenzierbar in einer Umgebung U von θ , und $M^{(4)}$ ist stetig in θ ,

(E) $aM'''(\theta) > \frac{1}{2} - 2\gamma$

erfüllt, dann gilt für $\gamma = \frac{1}{8}$:

(2) $n^{\frac{1}{8}}(X_{n+1} - \theta) = \mu(\frac{1}{8}) + S(\frac{1}{8}) + o(1)$ f.s. ,

für $\gamma > \frac{1}{8}$:

(3) $$n^{\frac{1}{2}-2\gamma}(X_{n+1} - \theta) = S(\gamma) + o(1) \quad \text{f.s.}$$

und für $\gamma < \frac{1}{8}$:

(4) $$n^{2\gamma}(X_{n+1} - \theta) \to \mu(\gamma) \quad \text{f.s.}$$

mit

$$\mu(\gamma) := -\frac{1}{12}(aM'''(\theta) - 2\gamma)^{-1}ac^2M^{(4)}(\theta)$$

$0 < \gamma \leq \frac{1}{8}$, und

$$S(\gamma) := -ac^{-2}n^{-aM'''(\theta)+\frac{1}{2}-2\gamma}\sum_{i=1}^{n} i^{aM'''(\theta)-1+2\gamma}U_i \quad .$$

Bemerkungen:

i) Die Wahl von $\gamma = \frac{1}{8}$ ist wieder in dem Sinne optimal, daß sich für größere γ sofort und für kleinere γ im Falle $M^{(4)}(\theta) \neq 0$ schlechtere Konvergenzraten ergeben. Im Falle $M^{(4)}(\theta) = 0$ und unter zusätzlichen Voraussetzungen an höhere Ableitungen läßt sich sicherlich in Analogie zu Satz (5.6) eine höhere Konvergenzgeschwindigkeit mit einer anderen Wahl von γ erzielen.

ii) Erfüllen die Fehler an Stelle von (C):
Es ex. $\sigma > 0$ und eine Brown'sche Bewegung B , so daß gilt:

$$\Sigma_{i \leq t} U_i = \sqrt{6}\,\sigma B(t) + o(t^{\frac{1}{2}}) \quad \text{f.s.} \quad ,$$

dann lassen sich mit Hilfe der beim Robbins-Monro- und Kiefer-Wolfowit Prozeß beschriebenen Methoden wieder konkretere Ergebnisse zeigen. (De Normierungsfaktor $\sqrt{6}$ erklärt sich daraus, daß sich die U_n aus Beobachtungsfehlern U_n^+, U_n^- bzw. U_n^0 von M an den Stellen $X_n + c_n$, $X_n - c_n$ bzw. X_n zusammensetzen: $U_n = U_n^+ + U_n^- - 2U_n^0$ (vgl. Kapitel 1 und die Bemerkung zu Korollar (5.5)).):

Für den A_3-Prozeß mit $\gamma = \frac{1}{8}$ gilt

a) ein fast sicheres Invarianzprinzip:

Es existiert eine Brown'sche Bewegung B', so daß gilt:

$$[t]^{\frac{1}{4}}(X_{[t]} - \theta) = \mu + \rho t^{-aM'''(\theta)} + {}^{\frac{1}{4}}B'(t^{2aM'''(\theta)} - \tfrac{1}{2}) + o(1) \quad \text{f.s.} \quad ;$$

b) ein Gesetz vom iterierten Logarithmus:

$$\limsup_{n\to\infty} \frac{n^{\frac{1}{4}}|X_n - \theta|}{\sqrt{2\ \text{loglog}n}} = \rho \quad \text{f.s.} \quad ;$$

c) asymptotische Normalität:

$$n^{\frac{1}{4}}(X_n - \theta) \xrightarrow{\mathcal{D}} Y \sim N(\mu, \rho^2) \quad .$$

Dabei ist jeweils

$$\mu = \mu(\tfrac{1}{8}) = -\ (12aM'''(\theta) - 3)^{-1} ac^2 M^{(4)}(\theta) \quad ,$$

$$\rho = a\sigma\sqrt{6}\ c^{-2}(2aM'''(\theta) - \tfrac{1}{2})^{-\frac{1}{2}} \quad .$$

Beweis von Satz (5.7):

Mit $A_n := an^{-1}$, $E_n := c_n^{-2}U_n$ und f_n definiert durch $f_n(x) = c_n^{-2}(M(x + c_n) + M(x - c_n) - 2M(x))$ besitzt der A_3-Prozeß die Form (2.3), und die Voraussetzungen (D) und (E) von Satz (2.4) sind erfüllt. Dank Taylor-Entwicklung existieren für alle x und alle n ein $\xi^+(x,n)$ und ein $\xi^-(x,n)$ mit $|\xi^\pm(x,n) - x| < c_n$ und

$$M(x \pm c_n) = M(x) \pm c_n M'(x) + \frac{1}{2} c_n^2 M''(\xi^\pm(x,n)) \quad .$$

Folglich gilt:

$$f_n(x) = \frac{1}{2}(M''(\xi^+(x,n)) + M''(\xi^-(x,n))) \quad .$$

Die Voraussetzungen (2.4) (A) bis (C) an die Funktionen f_n lassen sich nun durch die Eigenschaften von M'' genauso verifizieren, wie es im Be-

weis von (5.4) (1) durch die Eigenschaften der dortigen Funktion M' geschehen ist. Wir verzichten auf eine langweilige Wiederholung dieses Beweisschrittes und schließen mit Satz (2.4) auf die fast sichere Konvergenz von X_n gegen θ .

Wir setzen nun $f := M''$ und $E_n' := E_n + B_n$ mit $B_n := f_n(X_n) - f(X_n)$. Wegen $E_n = c^{-2}n^{2\gamma}U_n$ folgt aus (C)

$$\Sigma_{n=1}^{\infty} n^{-1+\delta}E_n \text{ konvergiert f.s. für alle } \delta < \frac{1}{2} - 2\gamma \quad .$$

Gelingt es uns noch,

$$(5) \qquad n^{2\gamma}B_n \to \frac{1}{12}\, c^2 M^{(4)}(\theta) \quad \text{f.s.}$$

unter der Voraussetzung (D) zu zeigen, dann konvergiert $\Sigma_{n=1}^{\infty} n^{-1+\delta}B_n$ f.s. für alle $\delta < 2\gamma$, und für $\Delta = \min(2\gamma, \frac{1}{2} - 2\gamma)$ können wir Satz (4.1) anwenden, so daß wir

$$n^{\Delta}(X_{n+1} - \theta) = -\, an^{-aM'''(\theta)+\Delta} \Sigma_{i=1}^{n} i^{aM'''(\theta)-1}B_i$$

$$- ac^{-2}n^{-aM'''(\theta)+\Delta} \Sigma_{i=1}^{n} i^{aM'''(\theta)-1+2\gamma}U_i + o(1)$$

erhalten. Unter Anwendung des Lemmas von Kronecker bzw. des Lemmas von Toeplitz ergeben sich wie im Beweis von Satz (5.4) die Behauptungen (2) bis (4).

Der Beweis von (5) verläuft wieder nach dem üblichen Schema über Realisationen ω mit $\lim_{n\to\infty} X_n(\omega) = \theta$, was ja nach (1) für fast alle ω erfüllt ist. Für ein derartiges ω gibt es nach (D) ein N , so daß M für $n \geq N$ auf dem Intervall $[x_n - c_n\,,\ x_n + c_n]$ viermal differenzierbar ist. Taylor-Entwicklung liefert:

$$M(x_n \pm c_n) = M(x_n) \pm c_n M'(x_n) + \frac{1}{2} c_n^2 M''(x_n) \pm \frac{1}{6} c_n^3 M'''$$

$$+ \frac{1}{24} c_n^4 M^{(4)}(\xi_n^{\pm})$$

für gewisse ξ_n^+, ξ_n^- mit $x_n - c_n < \xi_n^- < x_n < \xi_n^+ < x_n + c_n$. Damit ergibt sich:

$$b_n = c_n^{-2}(M(x_n + c_n) + M(x_n - c_n) - 2M(x_n)) - M''(x_n)$$

$$= \frac{1}{24} c_n^2 (M^{(4)}(\xi_n^+) + M^{(4)}(\xi_n^-)) \quad .$$

Unter Beachtung von $c_n^2 = c^2 n^{-2\gamma}$ und $\lim_{n\to\infty} \xi_n^+ = \lim_{n\to\infty} \xi_n^- = \theta$ folgt mit der Stetigkeit von $M^{(4)}$ die Aussage (5). □

Kapitel 6: Konvergenz zufällig gewichteter Summen

Im vorangegangenen Kapitel ließ sich auf Grund der deterministischen Schrittweiten $A_n = an^{-1}$ und der Struktur der betrachteten speziellen stochastischen Approximationsprozesse die Forderung der fast sicheren Konvergenz von $\Sigma_{n=1}^{\infty} n^{\delta} A_n E_n$ auf die f.s. Konvergenz deterministisch gewichteter Summen der zugrunde liegenden Störungen U_n zurückführen. Dieses war zum Beispiel dann erfüllt, wenn die U_n einem fast sicheren Invarianzprinzip genügen. Bevor wir nun adaptive stochastische Approximationsverfahren untersuchen können, müssen wir uns überlegen, ob die Forderung der f.s. Konvergenz von $\Sigma_{n=1}^{\infty} n^{\delta} A_n E_n$ sinnvoll ist. Es drängt sich also die Frage auf: Für welche zufälligen Störungen U_n und für welche zufälligen Gewichte W_n konvergiert $\Sigma_{n=1}^{\infty} W_n U_n$ fast sicher ?

Zur Beantwortung dieser Frage werden wir zuerst ein Ergebnis für Martingaldifferenzfolgen zitieren:

(6.1) Satz:

Ist $(U_n, A_n)_{n \in \mathbb{N}}$ eine Martingaldifferenzfolge mit $\sup_n E(U_n^2) < \infty$, und sind W_n A_{n-1}-meßbare Zufallsvariable, $n \in \mathbb{N}$, für die gilt:

(A) es existiert eine Folge $(\alpha_n)_{n \in \mathbb{N}}$ positiver reeller Zahlen mit $\Sigma_{n=1}^{\infty} \alpha_n^2 < \infty$ und $\limsup_{n \to \infty} \alpha_n^{-1} |W_n| < \infty$ f.s.,

dann konvergiert $\Sigma_{n=1}^{\infty} W_n U_n$ f.s.

Dieses Resultat ist eine einfache Anwendung von Satz 2.9.4 bei Stout (1974), S. 77, wobei wir Stout's S_n bzw. ν_n mit $\Sigma_{i=1}^{n} \alpha_i U_i$ bzw. $\alpha_n^{-1} W_n$ identifizieren müssen. $\sup_n E|S_n| < \infty$ gilt dann wegen

$(E|S_n|)^2 \leq E(S_n^2) = \Sigma_{i=1}^n \alpha_i^2 E(U_i^2)$ nach Voraussetzung.

Bilden die U_n keine Martingaldifferenzfolge, müssen wir eine eventuell vorhandene Struktur der Gewichte W_n ausnutzen.

Eine Zuwachsbedingung

Die Idee der weiteren Vorgehensweise ist nun folgende: Wir teilen die Indexmenge der natürlichen Zahlen in aufeinanderfolgende Blöcke ein und setzen voraus, daß die Gewichte W_n innerhalb der Blöcke "nicht zu stark" voneinander abweichen und daß die Zuwächse der Partialsummen $\Sigma_{i=m}^n U_i$ innerhalb eines Blocks "nicht viel größer" als die Quadratwurzel aus der Länge des Blocks sein sollen. Um diese Aussagen exakter zu machen, definieren wir für alle $\mu \geq 1$ die Blöcke

$$N(k,\mu) := \{n | k^\mu \leq n < (k+1)^\mu\} \quad ,$$

$k \in \mathbb{N}$, und bezeichnen mit $m(k,\mu)$ die kleinste natürliche Zahl aus $N(k,\mu)$. $m(k,\mu)$ ist wohldefiniert, da nach Mittelwertsatz $(k+1)^\mu - k^\mu \geq \mu k^{\mu-1} \geq 1$ für $\mu \geq 1$ und somit $N(k,\mu)$ nicht leer ist. Zwecks Schreinvereinfachung führen wir noch die vergrößerten Blöcke

$$N'(k,\mu) := N(k,\mu) \cup \{m(k+1,\mu)\}$$

ein. Wir können nun die zentrale Aussage über Konvergenz zufällig gewichteter Summen unter Zuwachsbedingung formulieren:

(6.2) Satz:

Genügen die Zufallsvariablen U_n, W_n und die reellen Zahlen $\tau > 0$, $\rho > 0$ und ψ mit $0 < \psi < \min(\rho, \frac{1}{2} + \tau)$ den Voraussetzungen

(A) $W_n = O(n^{-\frac{1}{2}-\tau})$ f.s. ,

(B) $\sup_{n \in N'(k,\mu)} |W_n - W_{m(k,\mu)}| = O(k^{-\frac{1}{2}\mu - \frac{1}{2} - \rho})$ f.s. für alle $\mu \geq 1$,

(C) $\sup_{n \in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} U_i| = O(k^{\frac{1}{2}\mu - \frac{1}{2} + \psi})$ f.s. für alle $\mu \geq 1$,

dann konvergiert $\Sigma_{n=1}^{\infty} W_n U_n$ f.s.

Bemerkung:

Die Zuwachsbedingung (C) ist aus dem Wunsch entstanden, eine Aussage von der Art eines Gesetzes vom iterierten Logarithmus zu haben. Eine derartige Aussage wäre etwa die Gültigkeit von (C) für alle $\psi > 0$. Dies ist aber bereits im Falle unabhängiger, identisch verteilter Zufallsvariablen U_n nur unter starken zusätzlichen Voraussetzungen möglich (vgl. Csörgö und Révész (1981), S. 117 f).

Beweis:

Wir werden eine Art pervertierter Induktion durchführen. Dazu werden wir als erstes zeigen ("Induktionsanfang"):

(1) $\Sigma_{k=1}^{\infty} W_{m(k,\mu)} \Sigma_{i \in N(k,\mu)} U_i$ konvergiert f.s. für alle $\mu > 1 + \frac{1}{\tau}$

Der entscheidende Teil wird aber der Beweis folgender Behauptung ("Induktionsschluß") sein:

(2) Für alle $\mu \geq 1$ und alle ν mit $1 \leq \nu < (1 + 2\psi)^{-1}(1 + 2\rho)$ gilt
$\Sigma_{j=1}^{\infty} W_{m(j,\mu)} \Sigma_{i \in N(j,\mu)} U_i$ konvergiert f.s. ,
falls $\Sigma_{k=1}^{\infty} W_{m(k,\nu\mu)} \Sigma_{i \in N(k,\nu\mu)} U_i$ f.s. konvergiert.

Wegen $\psi < \rho$ ist die Existenz eines geeigneten $\nu > 1$ gesichert. Zu diesem ν gibt es aber eine natürliche Zahl N mit $\nu^N > 1 + \frac{1}{\tau}$, so daß auf Grund von (1) die f.s. Konvergenz von $\Sigma_{k=1}^{\infty} W_{m(k,\nu^N)} \Sigma_{i\in N(k,\nu^N)} U_i$ gesichert ist. Unter Beachtung von $N(j,1) = \{j\}$ folgt nach N-maliger Anwendung von (2) die Behauptung des Satzes. Es bleiben noch (1) und (2) zu zeigen:

Unter Ausnutzung von $m(k,\mu) = O(k^{\mu})$ folgt aus (A) und (C)

$$W_{m(k,\mu)} \Sigma_{i\in N(k,\mu)} U_i = O(k^{-\tau\mu-\frac{1}{2}+\psi}) \quad \text{f.s.}$$

Nach Voraussetzung gilt aber $-\tau\mu - \frac{1}{2} + \psi < -\tau\mu + \tau$, so daß für $\mu > 1 + \frac{1}{\tau}$ der auf der rechten Seite auftretende Exponent kleiner als -1 ist. Dies impliziert die f.s. Konvergenz in (1).

Da die Blöcke $N(k,\nu)$ für festes ν eine Partition der natürlichen Zahlen sind, gibt es zu jeder natürlichen Zahl j genau ein k(j) mit $j \in N(k(j),\nu)$. Mit dieser Bezeichnungsweise gelten dann offensichtlich Relationen wie $m(k(j),\nu\mu) = O(j^{\mu})$, $k(j) = O(j^{1/\nu})$, $m(j,\mu) - m(k(j),\nu\mu) = O(j^{\mu-1/\nu})$, $k(j)^{-1} = O(j^{-1/\nu})$ usw. Wir wollen die Eigenschaft (B) der Gewichte ausnutzen und teilen dazu die zu untersuchende Summe aus (2) wie folgt auf.

(3)
$$\begin{aligned}
&\Sigma_{j=1}^{\infty} W_{m(j,\mu)} \Sigma_{i\in N(j,\mu)} U_i \\
&= \Sigma_{j=1}^{\infty} (W_{m(j,\mu)} - W_{m(k(j),\nu\mu)}) \Sigma_{i\in N(j,\mu)} U_i \\
&\quad + \Sigma_{j=1}^{\infty} W_{m(k(j),\nu\mu)} \Sigma_{i\in N(j,\mu)} U_i \quad .
\end{aligned}$$

Voraussetzung (B) und (C) liefern:

$$\begin{aligned}
&(W_{m(j,\mu)} - W_{m(k(j),\nu\mu)}) \Sigma_{i\in N(j,\mu)} U_i \\
&= O(k(j)^{-\frac{1}{2}\nu\mu-\frac{1}{2}+\rho} \, j^{\frac{1}{2}\mu-\frac{1}{2}+\psi}) \\
&= O(j^{-(\frac{1}{2}+\rho)/\nu-\frac{1}{2}+\psi}) \quad \text{f.s.}
\end{aligned}$$

Da nach Voraussetzung an ν der letzte Exponent kleiner als - 1 ist, konvergiert die erste Summe auf der rechten Seite von (3) f.s. Bei der zweiten Summe untersuchen wir die Partialsummen:

$$(4)\qquad \sum_{j=1}^{J} W_{m(k(j),\nu\mu)} \sum_{i\in N(j,\mu)} U_i = \sum_{k=1}^{k(J)-1} W_{m(k,\nu\mu)} \sum_{j\in N(k,\nu)} \sum_{i\in N(j,\mu)} U_i + R(J)$$

mit

$$R(J) = W_{m(k(J),\nu\mu)} \sum_{j=m(k(J),\nu)}^{J} \sum_{i\in N(j,\mu)} U_i \quad .$$

Da die $N(j,\mu)$ disjunkt sind und auf Grund einer recht langweiligen Überlegung $\bigcup_{j\in N(k,\nu)} N(j,\mu) = \{i \mid m(k,\nu)^{\mu} \le i < m(k+1,\nu)^{\mu}\}$ gilt, ergibt sich

$$(5)\qquad \sum_{j\in N(k,\nu)} \sum_{i\in N(j,\mu)} U_i = \sum_{i\in N(k,\nu\mu)} U_i + S(k+1) - S(k)$$

mit

$$S(k) = \sum_{k^{\nu\mu} \le i < m(k,\nu)^{\mu}} U_i \quad .$$

Dabei ist zu beachten, daß wegen $m(k,\nu) \ge k^{\nu}$ tatsächlich $m(k,\nu)^{\mu} \ge k^{\nu\mu}$ gilt. Wir erhalten folglich:

$$(6)\qquad \sum_{k=1}^{K} W_{m(k,\nu\mu)} \sum_{j\in N(k,\nu)} \sum_{i\in N(j,\mu)} U_i = \sum_{k=1}^{K} W_{m(k,\nu\mu)} \sum_{i\in N(k,\nu\mu)} U_i + R'(K)$$

mit

$$R'(K) = \sum_{k=1}^{K} W_{m(k,\nu\mu)} (S(k+1) - S(k))$$

$$= W_{m(K,\nu\mu)} S(K+1) - \sum_{k=1}^{K-1} (W_{m(k+1,\nu\mu)} - W_{m(k,\nu\mu)}) S(k+$$

$$- W_1 S(1) \quad .$$

Nun müssen wir die Restterme untersuchen:

Nach Definition von $m(k,\nu)$ gilt $m(k,\nu) < (k+1)^{\nu}$ und folglich $m(k,\nu)^{\mu} < (k+1)^{\nu\mu}$. Für $S(k)$ ergibt sich also nach Voraussetzung (C)

$$S(k) = O(k^{\frac{1}{2}\nu\mu - \frac{1}{2} + \psi}) \quad \text{f.s.}$$

Nach (A) und wegen $\psi < \frac{1}{2} + \tau$ gilt somit:

$$W_{m(K,\nu\mu)} S(K+1) = O(K^{-\tau\nu\mu - \frac{1}{2} + \psi}) = o(1) \quad \text{f.s.}$$

und nach (B)

$$(W_{m(k+1,\nu\mu)} - W_{m(k,\nu\mu)}) S(k+1) = O(k^{-1-\rho+\psi}) \quad \text{f.s.}$$

Wegen $\psi < \rho$ konvergiert die Summe über diese Terme f.s. Insgesamt konvergiert also $R'(K)$ f.s. gegen eine Zufallsvariable. Die in (2) vorausgesetzte f.s. Konvergenz von $\Sigma_{k=1}^{\infty} W_{m(k,\nu\mu)} \Sigma_{i \in N(k,\nu\mu)} U_i$ impliziert nun via (6), daß $\Sigma_{k=1}^{k(J)-1} W_{m(k,\nu\mu)} \Sigma_{j \in N(k,\nu)} \Sigma_{i \in N(i,\mu)} U_i$ (für $J \to \infty$) f.s. gegen eine Zufallsvariable konvergiert. Gelingt es uns noch,

$$R(J) \to 0 \quad \text{f.s.}$$

zu zeigen, dann folgt aus (4) die Behauptung (2). Mit denselben Überlegungen wie zu (5):

$\cup_{j=m(k(J),\nu)}^{J} N(j,\mu) = \{i \mid m(k(J),\nu)^{\mu} \leq i < (J+1)^{\mu}\}$ ergibt sich:

$$\Sigma_{j=m(k(J),\nu)}^{J} \Sigma_{i \in N(j,\mu)} U_i$$

$$= \Sigma_{m(k(J),\nu\mu) \leq i < (J+1)^{\mu}} U_i - S(k(J)) \quad .$$

Nach Definition von $k(J)$ gilt $J + 1 \leq m(k(J)+1,\nu) < (k(J)+2)^{\nu}$, und durch kreative Anwendung der Voraussetzung (C) (der O-Term ändert sich nicht, wenn wir das Supremum nicht nur über $N(k,\mu)$, sondern zusätzlich auch über $N(k+1,\mu)$ betrachten) erhalten wir

$$\Sigma_{j=m(k(J),\nu)}^{J} \Sigma_{i \in N(j,\mu)} U_i = O(k(J)^{\frac{1}{2}\nu\mu - \frac{1}{2} + \psi}) \quad \text{f.s.}$$

Damit ergibt sich unter Verwendung von (A):

$$R(J) = O(k(J)^{-\tau\nu\mu - \frac{1}{2} + \psi}) \quad \text{f.s.} \quad ,$$

woraus wegen $\psi < \frac{1}{2} + \tau$ das erhoffte Ergebnis $R(J) = o(1)$ f.s. folgt. □

Nach diesem recht langwierigen Beweis wollen wir uns bei einer Anwendung auf deterministische Gewichte entspannen:

(6.3) Korollar:

Genügen die Zufallsvariablen U_n und $\psi > 0$ der Voraussetzung (6.2) (C dann konvergiert $\Sigma_{n=1}^{\infty} n^{-\frac{1}{2}-\tau} U_n$ f.s. für alle $\tau > \max(\psi - \frac{1}{2}, 0)$.

Beweis:

Wir setzen $W_n = n^{-\frac{1}{2}-\tau}$. Da für alle $n \in N'(k,\mu)$

$$0 \leq m(k,\mu)^{-\frac{1}{2}-\tau} - n^{-\frac{1}{2}-\tau} \leq k^{-\frac{1}{2}\mu-\tau\mu} - (k+2)^{-\frac{1}{2}\mu-\tau\mu}$$

gilt, folgt mit der üblichen Mittelwertsatzargumentation

$$\sup_{n \in N'(k,\mu)} |n^{-\frac{1}{2}-\tau} - m(k,\mu)^{-\frac{1}{2}-\tau}| = O(k^{-\frac{1}{2}\mu-\tau\mu-1}) \quad .$$

Somit ist (6.2) (B) mit $\rho = \frac{1}{2} + \tau$ erfüllt. Nach Voraussetzung gilt $\psi < \frac{1}{2} + \tau$, und (6.2) (A) ist trivialerweise erfüllt, so daß uns Satz (6.2) das gewünschte Ergebnis liefert. □

Diese Aussage läßt unsere Zuwachsbedingung in der Hinsicht sinnvoll erscheinen, daß sie mit der bisherigen Forderung der Konvergenz deterministisch gewichteter Summen verträglich ist. Außerdem gibt sie uns die Möglichkeit, weitere Klassen von zufälligen Störungen zu betrachten die z.B. nicht einem starken Invarianzprinzip der Form (5.1) genügen. In der Anwendung werden wir noch eine Abschätzung für die Zuwächse deterministisch gewichteter Summen benötigen. Da es sich um eine rein

analytische Aussage handelt, werden wir sie auch als solche formulieren:

(6.4) Lemma:

Genügt eine Folge $(u_n)_{n\in\mathbb{N}}$ reeller Zahlen der Voraussetzung

(A) $\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} u_i| = O(c_k)$

für ein $\mu \geq 1$ und eine Folge $(c_k)_{k\in\mathbb{N}}$ positiver reeller Zahlen, dann gilt für alle $\alpha \in \mathbb{R}$:

$$\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} i^{\alpha} u_i| = O(k^{\alpha\mu} c_k) \quad .$$

Beweis:

Wir werden ausnutzen, daß wir i^{α} für $i \in N(k,\mu)$ wie folgt in einer Teleskopsumme darstellen können:

$$i^{\alpha} - m(k,\mu)^{\alpha} = \Sigma_{j=m(k,\mu)+1}^{i} (j^{\alpha} - (j-1)^{\alpha}) \quad .$$

Da die Summanden in der Teleskopsumme alle das gleiche Vorzeichen besitzen, gilt sogar:

$$|i^{\alpha} - m(k,\mu)^{\alpha}| = \Sigma_{j=m(k,\mu)+1}^{i} |j^{\alpha} - (j-1)^{\alpha}| \quad .$$

Wir erhalten also:

$$\Sigma_{i=m(k,\mu)}^{n} i^{\alpha} u_i = \Sigma_{i=m(k,\mu)}^{n} \Sigma_{j=m(k,\mu)+1}^{i} (j^{\alpha} - (j-1)^{\alpha}) u_i + m(k,\mu)^{\alpha} \Sigma_{i=m(k,\mu)}^{n} u_i$$

und nach Umsummieren der Doppelsumme:

$$\Sigma_{i=m(k,\mu)}^{n} i^{\alpha} u_i = \Sigma_{j=m(k,\mu)+1}^{n} (j^{\alpha} - (j-1)^{\alpha}) \Sigma_{i=j}^{n} u_i + m(k,\mu)^{\alpha} \Sigma_{i=m(k,\mu)}^{n} u_i \quad .$$

Nun gibt es nach (A) eine Konstante C , so daß $|\Sigma_{i=j}^{n} u_i| \le Cc_k$ für alle $j,n \in N(k,\mu)$, $j \le n$, gilt. Somit ergibt sich

$$|\Sigma_{i=m(k,\mu)}^{n} i^{\alpha} u_i| \le \Sigma_{j=m(k,\mu)+1}^{n} |j^{\alpha} - (j-1)^{\alpha}| Cc_k + m(k,\mu)^{\alpha} Cc_k$$

$$= (|n^{\alpha} - m(k,\mu)^{\alpha}| + m(k,\mu)^{\alpha}) Cc_k \quad ,$$

und folglich:

$$\sup_{n \in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} i^{\alpha} u_i| \le (m(k+1,\mu)^{\alpha} + 2m(k,\mu)^{\alpha}) Cc_k \quad ,$$

und die rechte Seite ist von der Größenordnung $O(k^{\alpha\mu} c_k)$. □

Eine einfache Folgerung aus diesem Lemma ist, daß unter Voraussetzung (6.2) (C) für alle τ und für alle $\mu \ge 1$

$$\sup_{n \in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} i^{-\frac{1}{2}-\tau} U_i| = O(k^{-\tau\mu-\frac{1}{2}+\psi}) \quad \text{f.s.}$$

gilt. Mit diesem Ergebnis können wir Korollar (6.3) direkt und für alle $\tau > 0$, d.h. ohne Nebenbedingung $\tau > \psi - \frac{1}{2}$, beweisen:

Für $\mu \ge \max(\tau^{-1}(\psi + \frac{3}{2}), 1)$ gilt für den Exponenten in der Abschätzun $-\tau\mu - \frac{1}{2} + \psi \le -2$. Für dieses μ sei $j(n)$ wieder so definiert, daß $n \in N(j(n),\mu)$ erfüllt ist. Mit dieser Schreibweise ergibt sich für alle $n \ge m(k,\mu)$:

$$|\Sigma_{i=m(k,\mu)}^{n} i^{-\frac{1}{2}-\tau} U_i| \le \Sigma_{j=k}^{j(n)-1} |\Sigma_{i \in N(j,\mu)} i^{-\frac{1}{2}-\tau} U_i|$$

$$+ |\Sigma_{i=m(j(n),\mu)}^{n} i^{-\frac{1}{2}-\tau} U_i| \quad .$$

Nutzen wir jetzt unsere Abschätzung aus, dann erhalten wir:

$$\sup_{n \ge m(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} i^{-\frac{1}{2}-\tau} U_i| = O(\Sigma_{j=k}^{\infty} j^{-2}) = O(k^{-1}) \quad \text{f.s.}$$

Somit bilden die Partialsummen f.s. eine Cauchy-Folge, woraus die f.s. Konvergenz der Summe $\Sigma_{n=1}^{\infty} n^{-\frac{1}{2}-\tau} U_n$ folgt.

Eine Momentenbedingung

Die Zuwachsbedingung (6.2)(C) ist sehr implizit, und wir wollen sie daher durch eine leichter nachprüfbare Bedingung an die Momente von Teilsummen ersetzen:

(6.5) Satz:

Genügen die Zufallsvariablen U_n und $p \geq 2$ der Voraussetzung

(A) für alle $\varepsilon > 0$ ex. $C_\varepsilon < \infty$ mit

$$E(|\Sigma_{i=a+1}^{a+m} U_i|^p) \leq C_\varepsilon m^{\frac{1}{2}p+\varepsilon} \quad \text{für alle } a,m \in \mathbb{N} ,$$

dann gilt für alle $\psi > p^{-1}$ und alle $\mu \geq 1$

$$\sup_{n \in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} U_i| = O(k^{\frac{1}{2}\mu-\frac{1}{2}+\psi}) \quad \text{f.s.}$$

Bemerkungen:

i) Voraussetzung (A) ist zum Beispiel für Martingaldifferenzfolgen mit gleichmäßig beschränkten p-ten Momenten $\sup_n E(|U_n|^p) < \infty$ sogar für $\varepsilon = 0$ erfüllt:

Nach Burkholder's Martingalversion der Marcinkiewicz-Zygmund-Ungleichung (vgl. Stout (1974), S. 149, oder Hall und Heyde (1980), S. 23) gibt es für $p \geq 2$ ein $C = C(p) < \infty$ mit

$$E(|\Sigma_{i=a+1}^{a+m} U_i|^p) \leq CE(|\Sigma_{i=a+1}^{a+m} U_i^2|^{\frac{1}{2}p})$$

für alle $a,m \in \mathbb{N}$. Wir wenden nun noch die Hölder-Ungleichung für Summen an (vgl. Stout (1974), S. 154):

$$\Sigma_{i=a+1}^{a+m} 1 \cdot U_i^2 \le (\Sigma_{i=a+1}^{a+m} 1)^{1-2/p} (\Sigma_{i=a+1}^{a+m} |U_i|^p)^{2/p}$$

$$= m^{1-2/p} (\Sigma_{i=a+1}^{a+m} |U_i|^p)^{2/p} \quad ,$$

und erhalten somit:

$$E(|\Sigma_{i=a+1}^{a+m} U_i|^p) \le C m^{\frac{1}{2}p-1} E(\Sigma_{i=a+1}^{a+m} |U_i|^p)$$

$$\le C m^{\frac{1}{2}p} \sup_n E(|U_n|^p) \quad .$$

ii) Der Faktor m^ε auf der rechten Seite der Abschätzung in (A) ist aus dem Grund eingeführt, um eine große Klasse von schwach abhängigen Zufallsvariablen zulassen zu können. Im dritten Teil des Anhangs werden wir für φ- und ρ-mischende (und andeutungsweise auch für α-mischende) Folgen von Zufallsvariablen mit gleichmäßig beschränkten p-ten Momenten Abschätzungen der Form:

$$E(|\Sigma_{i=a+1}^{a+m} U_i|^p) \le C m^{\frac{1}{2}p} L(m) \quad ,$$

für alle $a,m \in \mathbb{N}$, angeben, wobei L eine Funktion, wie etwa der Logarithmus mit $L(m) = O(m^\varepsilon)$ für alle ε ist. Daraus folgt dann sofort, daß (A) erfüllt ist. Speziell wird L oft eine langsam variierende Funktion sein (zu Definition und Eigenschaften langsam variierender Funktionen vgl. Loève (1977) I, S. 354 ff).

iii) Die Voraussetzung (A) kann noch etwas abgeschwächt werden, ohne daß der Beweis geändert werden muß, indem man eine geringfügige zusätzliche Abhängigkeit von a zuläßt, etwa in der Form:

$$E(|\Sigma_{i=a+1}^{a+m} U_i|^p) \le C_\varepsilon m^{\frac{1}{2}p} (a+m)^\varepsilon \quad .$$

Damit könnte eventuell in der Anwendung die Forderung der gleichmäßigen Beschränktheit der p-ten Momente durch $E(|U_n|^p) = O(n^\varepsilon)$, für alle $\varepsilon > 0$, ersetzt werden. Dies scheint mir jedoch keine für unsere Zwecke sinnvolle Verallgemeinerung zu sein.

iv) Vergleichen wir nun die Momentenbedingung (A) mit der Zuwachsbedingung (6.2) (C), so stellen wir fest, daß (A) im allgemeinen nicht viel strenger ist:

Sind die U_n unabhängige, identisch verteilte Zufallsvariable, dann ist (A) nach Bemerkung i) offensichtlich genau dann erfüllt, wenn $E(|U_n|^p) < \infty$ gilt. Nach dem Momenten-Lemma (vgl. Loève (1977) I, S. 254) ist dieses äquivalent zu

$$\Sigma_{n=1}^{\infty} P(|U_n| \geq n^{1/p}) = \Sigma_{n=1}^{\infty} P(|U_1| \geq n^{1/p}) < \infty \quad ,$$

und nach dem bekannten Lemma von Borel-Cantelli ist dieses wiederum äquivalent zu

$$P(|U_n| \geq n^{1/p} \text{ unendlich oft}) = 0 \quad .$$

Daraus folgt $U_n = O(n^{1/p})$ f.s., d.h. (6.2) (C) für $\mu = 1$ mit $\psi = p^{-1}$. Ist umgekehrt (6.2) (C) für ein $\psi < p^{-1}$ erfüllt, d.h. speziell $U_n = O(n^\psi)$ f.s. für $\mu = 1$, dann gilt die Rückrichtung.

Beweis:

Da $m^{1+2\varepsilon/p} + n^{1+2\varepsilon/p} \leq (m+n)^{1+2\varepsilon/p}$ gilt, können wir Serfling's (1970) Verallgemeinerung der Rademacher-Menshov-Ungleichung (vgl. auch Billingsley (1968), S. 94, S. 102) auf unsere Voraussetzung anwenden, und wir erhalten:

$$E\Big(\sup_{1 \leq n \leq m} |\Sigma_{i=a+1}^{a+n} U_i|^p\Big) \leq C_\varepsilon (\log_2 2m)^{p+2\varepsilon} m^{\frac{1}{2}p+\varepsilon}$$

für alle $a,m \in \mathbb{N}$ und alle $\varepsilon > 0$, wobei $\log_2$ der Logarithmus zur Basis ist. Nun ist aber $(\log_2 2m)^{p+2\varepsilon}$ von der Größenordnung $O(m^{\varepsilon})$, und somit existiert für alle $\varepsilon > 0$ ein $C'_\varepsilon < \infty$ mit

$$E\Big(\sup_{1\le n\le m} |\Sigma_{i=a+a}^{a+n} U_i|^p\Big) \le C'_\varepsilon m^{\frac{1}{2}p + 2\varepsilon}$$

für alle $a,m \in \mathbb{N}$. Unter Anwendung der Markov-Ungleichung ergibt sich damit für alle $\varepsilon > 0$:

$$\begin{aligned} &P\Big(\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} U_i| > k^{\frac{1}{2}\mu - \frac{1}{2} + \psi}\Big) \\ &\le k^{-(\frac{1}{2}\mu - \frac{1}{2} + \psi)p} E\Big(\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} U_i|^p\Big) \\ &\le C'_\varepsilon k^{-\frac{1}{2}\mu p + \frac{1}{2}p - \psi p} (m(k+1,\mu) - m(k,\mu))^{\frac{1}{2}p + 2\varepsilon} \quad . \end{aligned}$$

Wegen $m(k+1,\mu) - m(k,\mu) = O(k^{\mu-1})$ folgt daraus

$$P\Big(\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} U_i| > k^{\frac{1}{2}\mu - \frac{1}{2} + \psi}\Big) = O(k^{2(\mu-1)\varepsilon - \psi p})$$

für alle $\varepsilon > 0$. Auf Grund von $\psi p > 1$ gibt es ein geeignetes $\varepsilon > 0$, so daß der Exponent $2(\mu - 1)\varepsilon - \psi p$ kleiner als -1 wird. Das bedeutet aber, daß die Summe über die Wahrscheinlichkeiten konvergiert, und nach dem Lemma von Borel-Cantelli gilt

$$P\Big(\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} U_i| > k^{\frac{1}{2}\mu - \frac{1}{2} + \psi} \text{ unendlich oft}\Big) = 0 \quad ,$$

woraus die Behauptung folgt. □

Für Martingaldifferenzfolgen mit $\sup_n E(U_n^2) < \infty$ konvergiert nach Satz (6.1) für alle $\delta < \frac{1}{2}$ die deterministisch gewichtete Summe $\Sigma_{n=1}^{\infty} n^{-1+\delta} U_n$ f.s. Zum Abschluß dieses Kapitels können wir aber nun als triviale Folgerung unserer bisherigen Ergebnisse eine derartige Aussage auch für eine größere Klasse von Folgen von Zufallsvariablen formulieren:

(6.6) Korollar:

Genügen die Zufallsvariablen U_n der Voraussetzung

(A) für alle $\varepsilon > 0$ ex. $C_\varepsilon < \infty$ mit

$$E(|\Sigma_{i=a+1}^{a+m} U_i|^2) \leq C_\varepsilon m^{1+\varepsilon} \quad \text{für alle } a,m \in \mathbb{N} ,$$

dann konvergiert $\Sigma_{n=1}^{\infty} n^{-1+\delta} U_n$ f.s. für alle $\delta < \frac{1}{2}$.

Beweis:

Nach Satz (6.5) erfüllen die Zufallsvariablen U_n die Voraussetzung (6.2) (C) für alle $\psi > \frac{1}{2}$, und mit Korollar (6.3) folgt die Behauptung. □

Kapitel 7: Der Venter-Prozeß

Mit Hilfe der bisherigen Ergebnisse können wir nun für den in der Einleitung motivierten adaptiven Venter-Prozeß Konvergenzaussagen und eine starke Repräsentation formulieren. Dazu fordern wir, daß die auftretenden zufälligen Störungen U_n und V_n einer von den beiden folgenden Bedingungen genügen:

(7.1) Martingalbedingung:

$(U_n, A_n)_{n\in\mathbb{N}}$ und $(V_n, A_n)_{n\in\mathbb{N}}$ sind Martingaldifferenzfolgen mit einer geeigneten Folge $(A_n)_{n\in\mathbb{N}}$ von σ-Algebren und es gilt:

$$\sup_{n\in\mathbb{N}} E(U_n^2) < \infty \ , \ \sup_{n\in\mathbb{N}} E(V_n^2) < \infty \ .$$

(7.2) Zuwachsbedingung:

Es ex. $\psi > 0$, so daß für alle $\mu \geq 1$ gilt:

$$\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} U_i| = O(k^{\frac{1}{2}\mu - \frac{1}{2} + \psi}) \quad \text{f.s.} \ ,$$

$$\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)}^{n} V_i| = O(k^{\frac{1}{2}\mu - \frac{1}{2} + \psi}) \quad \text{f.s.}$$

Da die Störungen U_n und V_n sich linear aus U_n^+ und U_n^- zusammensetzen: $U_n = \frac{1}{2}(U_n^+ + U_n^-)$, $V_n = U_n^+ - U_n^-$ (vgl. Einleitung), ist es nur natürlich, an U_n und V_n dieselben Bedingungen zu stellen, wie es in (7.1) und (7.2) geschehen ist. Die Martingalbedingung umfaßt wieder den Fall bedingt unabhängiger Störungen mit gleichmäßig beschränkter Varianz der F_x . Dabei können die A_n etwa die von X_1 , $U_1,\ldots,U_n$, $V_1,\ldots,V_n$ erzeugten σ-Algebren sein. Die Zuwachsbedingung erlaubt uns neben Martingal-

differenzfolgen eine weitere Klasse von Fehlerfolgen zuzulassen.

Um einen möglichst übersichtlichen Satz zu formulieren, setzen wir $\psi = 0$, falls U_n und V_n die Bedingung (7.1) erfüllen:

(7.3) Satz:

Genügen $M: \mathbb{R} \to \mathbb{R}$ und $\theta \in \mathbb{R}$ den Voraussetzungen

(A) $M(x)(x - \theta) > 0$ für alle $x \neq \theta$,

(B) M ist Lipschitz-stetig ,

(C) M ist differenzierbar in einer Umgebung von θ und M' ist stetig in θ ,

und erfüllen Zufallsvariablen U_n und V_n die Bedingung (7.1) - und die Zufallsvariable X_1 ist A_1-meßbar - oder (7.2) mit $0 \leq \psi < \frac{1}{4}$, dann gilt für den durch

$$X_{n+1} = X_n - A_n Y_n$$

mit

$$Y_n = \frac{1}{2}(M(X_n + c_n) + M(X_n - c_n)) + U_n \ ,$$

$$D_n = n^{-1} \Sigma_{i=1}^{n} (2c_i)^{-1} (M(X_i + c_i) - M(X_i - c_i) + V_i) \ ,$$

$c_n = cn^{-\gamma}$, $c > 0$, $0 < \gamma < \frac{1}{2} - 2\psi$, und

$$A_n = \begin{cases} (n \log n)^{-1} & \text{für } D_n > \log n \\ n^{-1+\alpha} & \text{für } D_n < n^{-\alpha} \\ (nD_n)^{-1} & \text{sonst} \end{cases}$$

für $n \geq 3$, $A_1 = A_2 = 1$, $0 < \alpha \leq \frac{1}{4}$, definierten Venter-Prozeß:

(1) $X_n \to \theta$ f.s. ,

(2) $D_n \to M'(\theta)$ f.s.

Ist zusätzlich

(D) M ist zweimal differenzierbar in einer Umgebung von θ , $M'(\theta) > 0$ und M'' ist stetig in θ ,

(E) $\frac{1}{4} < \gamma < \frac{1}{2} - 2\psi$

erfüllt, dann gilt:

(3) $$n^{\frac{1}{2}}(X_{n+1} - \theta) = -M'(\theta)^{-1} n^{-\frac{1}{2}} \Sigma_{i=1}^{n} U_i + o(1) \quad \text{f.s.} ,$$

(4) $$n^{\frac{1}{2}-\gamma}(D_n - M'(\theta)) = (2c)^{-1} n^{-\frac{1}{2}-\gamma} \Sigma_{i=1}^{n} i^{\gamma} V_i + o(1) \quad \text{f.s.}$$

Bemerkungen:

i) Nach Satz (6.1) bzw. Korollar (6.3) konvergieren unter der Bedingung (7.1) bzw. (7.2) mit $\psi < \frac{1}{4}$ die Reihen $\Sigma_{n=1}^{\infty} n^{-1+\delta} U_n$ und $\Sigma_{n=1}^{\infty} n^{-1+\delta} V_n$ f.s. für alle $\delta < \frac{1}{2}$. Für die Konvergenzgeschwindigkeit von X_n bzw. D_n ergibt sich somit aus (3) bzw. (4) mit dem Lemma von Kronecker:

(5) $$n^{\delta}(X_n - \theta) \to 0 \quad \text{f.s. für alle } \delta < \frac{1}{2} ,$$

(6) $$n^{\tau}(D_n - M'(\theta)) \to 0 \quad \text{f.s. für alle } \tau < \frac{1}{2} - \gamma .$$

Dabei muß man bei der Herleitung von (5) wieder beachten, daß $(n+1)^{\delta} - n^{\delta} = O(n^{\delta-1})$ gilt, und (1) ausnutzen.

ii) Die oberen bzw. unteren Schranken $\log n$ bzw. $n^{-\alpha}$, an denen die D_n abgeschnitten werden, können allgemeiner als β_n bzw. α_n mit $\beta_n \geq \alpha_n > 0$ und $\lim_{n\to\infty} \beta_n = \infty$ und $\Sigma_{n=1}^{\infty} (n\beta_n)^{-1} = \infty$ bzw. $\lim_{n\to\infty} \alpha_n = 0$ und $\alpha_n^{-1} = O(n^{\frac{1}{4}+\varepsilon})$ für alle $\varepsilon > 0$ gewählt werden, ohne daß sich an den Konvergenzeigenschaften des Prozesses etwas ändert.

iii) Die Bedingung $\psi < \frac{1}{4}$ ist eingeführt worden, damit überhaupt geeignete γ mit $0 < \gamma < \frac{1}{2} - 2\psi$ existieren. Die Voraussetzung (E) impliziert sogar $\psi < \frac{1}{8}$. Für die Martingalbedingung (mit $\psi = 0$) ist dieses ohne Auswirkung. Wollen wir jedoch die Zuwachsbedingung durch die Momentenbedingung (vgl. Satz (6.5)) nachweisen, so benötigen wir die Existenz (und gleichmäßige Beschränktheit) der p-ten Momente $E(|U_n|^p)$ bzw. $E(|V_n|^p)$ für ein $p > 4$ für die Aussagen (1) und (2) und für ein $p > 8$ für die Aussagen (3) und (4). Genauer muß sogar $p > 2(\frac{1}{2} - \gamma)^{-1}$ erfüllt sein, so daß es sinnvoll ist, γ möglichst klein, d.h. möglichst nahe an $\frac{1}{4}$ zu wählen. Für ein derartiges γ ist auch die Konvergenzgeschwindigkeit von D_n möglichst gut (vgl. (4) und (6)). In Anschluß an den Beweis des Satzes werden wir uns überlegen, daß sich für $\gamma \leq \frac{1}{4}$ schlechtere Konvergenzeigenschaften ergeben. Es sei noch bemerkt, daß bei vielen praktischen Problemstellungen wie zum Beispiel das Auffinden des Quantils einer unbekannten Verteilungsfunktion die Fehler fast sicher gleichmäßig beschränkt sind, so daß die Existenz und gleichmäßige Beschränktheit der p-ten Momente für alle p gesichert sind.

iv) Fabian (1968) betrachtete nicht den hier vorliegenden, sondern den durch die Rekursion

$$X_{n+1} = X_n - A_{n-1}Y_n$$

definierten Prozeß mit ansonsten denselben Bezeichnungsweisen. Er verzichtete also bei der Näherung für $M'(\theta)^{-1}$ auf die jeweils letzte Beobachtung, um etwaige Störungen durch die Abhängigkeit von U_n und V_n zu vermeiden. Aus dem folgenden Beweis wird ersichtlich, daß wir unter der Martingalbedingung (7.1) für einen derartigen Prozeß wie Fabian

(1968) für α auch größere Werte: $0 < \alpha < \frac{1}{2}$ zulassen und von der Forderung (B) der Lipschitz-Stetigkeit wieder zu den üblichen Voraussetzungen

(B_1) $\inf_{\delta_1 \le |x-\theta| \le \delta_2} |M(x)| > 0$ für alle $\delta_2 \ge \delta_1 > 0$,

(B_2) es ex. $\alpha', \beta > 0$ mit $|M(x)| \le \alpha' + \beta|x - \theta|$ für alle x

übergehen können. Tatsächlich sind (B_1) und (B_2) schwächer als (B), da (B_1) sowie $M(\theta) = 0$, wie bereits schon einmal bemerkt, aus (A) und der Stetigkeit folgt und somit (B_2) für alle $\alpha' > 0$ erfüllt ist, wenn β die Lipschitz-Konstante ist.

Beweis:

Wir werden den Beweis für die beiden Bedingungen (7.1) und (7.2) an die Störungen gleichzeitig durchführen und nur an den kritischen Stellen, die die fast sichere Konvergenz gewisser gewichteter Summen betreffen, gesonderte Betrachtungen anstellen. Der Beweis ist recht langwierig - wenn nicht gar langweilig - und er ist daher zur besseren Orientierung in fünf Schritte eingeteilt, wobei die ungeradzahligen Aussagen über X_n und die geradzahligen Aussagen über D_n betreffen.

Schritt 1:

Mit f_n definiert durch $f_n(x) := \frac{1}{2}(M(x + c_n) + M(x - c_n))$ und $E_n := U_n$ besitzt der Venter-Prozeß die Form (2.3). Der Nachweis, daß die Voraussetzungen (2.4) (A) bis (C) an die f_n aus den entsprechenden an die Funktion M folgen, verläuft nach dem im Kapitel 5 beschriebenen Schema, das wir hier der Vollständigkeit halber noch einmal darlegen wollen:

Zu $\varepsilon > 0$ gibt es ein $N(\varepsilon)$, so daß $c_n < \varepsilon$ für $n \geq N(\varepsilon)$ gilt. Somit besitzen x mit $|x - \theta| > \varepsilon$ und $x + c_n$ sowie $x - c_n$ für $n \geq N(\varepsilon)$ dasselbe Vorzeichen, so daß (2.4) (A) aus (A) folgt. Ist $\varepsilon < \delta_1$, dann gilt für $n \geq N(\varepsilon)$

$$\inf_{\delta_1 \leq |x-\theta| \leq \delta_2} |f_n(x)| \geq \inf_{\delta_1 - \varepsilon \leq |x-\theta| \leq \delta_2 + \varepsilon} |M(x)| \quad ,$$

und (2.4) (B) ergibt sich aus (A) und (B) über (B_1). Schließlich folgt aus (A) und (B) über (B_2)

$$|f_n(x)| \leq \tfrac{1}{2}\beta(|x + c_n - \theta| + |x - c_n - \theta|) \leq \beta c_n + \beta|x - \theta| \quad ,$$

wobei β hier und im weiteren die Lipschitz-Konstante ist. Damit ist auch (2.4) (C) mit βc an Stelle von α erfüllt.

Nach Definition von A_n gilt für $n \geq 3$

$$(n \log n)^{-1} \leq A_n \leq n^{\alpha-1} \quad .$$

Damit erhalten wir einerseits

$$0 \leq \limsup_{n\to\infty} A_n \leq \lim_{n\to\infty} n^{\alpha-1} = 0 \quad ,$$

folglich:

$$\lim_{n\to\infty} A_n = 0 \quad ,$$

und andererseits mit Hilfe der Integralabschätzung für Summen (A.1)

$$\Sigma_{n=1}^{N} A_n \geq \Sigma_{n=3}^{N} (n \log n)^{-1} \geq \int_3^N (x \log x)^{-1} dx \geq \log\log N \quad .$$

Da $\log\log N$ über alle Grenzen wächst, folgt daraus

$$\Sigma_{n=1}^{\infty} A_n = \lim_{N\to\infty} \Sigma_{n=1}^{N} A_n = \infty \quad ,$$

und (2.4) (D) ist erfüllt.

Können wir jetzt noch

(7) $\Sigma_{n=1}^{\infty} A_n U_n$ konvergiert f.s.

zeigen, dann liefert uns Satz (2.4) die Aussage (1). Dazu stellen wir zuerst einige allgemeine Betrachtungen über Differenzen der Form

(8) $A_n - A_m = A_n A_m (A_m^{-1} - A_n^{-1})$

für $3 \leq m < n$ an. Die Definition von A_k , $k \geq 3$, liefert uns die Abschätzung

$$k^{1-\alpha} \leq A_k^{-1} \leq k \log k$$

mit $A_k^{-1} = k \log k$ für $D_k \geq \log k$, $A_k^{-1} = k^{1-\alpha}$ für $D_k \leq k^{-\alpha}$ und strikte Ungleichheit sonst. Wir machen jetzt einige elementare Fallunterscheidungen:

1. Fall: "$A_m^{-1} \geq A_n^{-1}$":

   Dann gilt $A_n^{-1} < m \log m < n \log n$

   bzw. $A_m^{-1} \geq n^{1-\alpha} > m^{1-\alpha}$ wegen $\alpha < 1$,

   was gleichbedeutend mit $A_n^{-1} \geq nD_n$ bzw. $A_m^{-1} \leq mD_m$

   ist, und damit:

$$|A_m^{-1} - A_n^{-1}| = A_m^{-1} - A_n^{-1} \leq mD_m - nD_n \quad .$$

2. Fall: "$A_m^{-1} < A_n^{-1}$":

   Es gilt: $|A_m^{-1} - A_n^{-1}| = A_n^{-1} - A_m^{-1}$.

   Unterfall 2.1: "$m^{-1}A_m^{-1} \geq n^{-1}A_n^{-1}$":

$$A_n^{-1} - A_m^{-1} \leq nm^{-1}A_m^{-1} - A_m^{-1} \leq n \log m - m \log m < n \log n - m \log m$$

Unterfall 2.2: "$m^{-1}A_m^{-1} < n^{-1}A_n^{-1}$":

A) "$A_m^{-1} \geq mD_m$":

Dann gilt $A_n^{-1} > nm^{-1}A_m^{-1} \geq nm^{-\alpha} > n^{1-\alpha}$, was $A_n^{-1} \leq nD_n$ bedeutet, folglich

$$A_n^{-1} - A_m^{-1} \leq nD_n - mD_m \quad .$$

B) "$A_m^{-1} < mD_m$":

Dieses liefert $A_m^{-1} = m \log m$ und somit

$$A_n^{-1} - A_m^{-1} \leq n \log n - m \log m \quad .$$

Als Fazit dieses Exkurses in die Niederungen der Elementarmathematik erhalten wir:

$$(9) \qquad |A_m^{-1} - A_n^{-1}| \leq |nD_n - mD_m| + (n \log n - m \log m)$$

für $3 \leq m < n$, und für $m = n$ ist (9) trivialerweise richtig. Jetzt erinnern wir uns der Struktur der D_k:

$$(10) \qquad nD_n - mD_m = \Sigma_{i=m+1}^{n} (2c_i)^{-1} (M(X_i + c_i) - M(X_i - c_i) + V_i) \quad .$$

Nutzen wir nun die Lipschitz-Stetigkeit (B) aus, so bekommen wir:

$$(2c_i)^{-1} |M(X_i + c_i) - M(X_i - c_i)| \leq B \quad ,$$

und folglich:

$$|nD_n - mD_m| \leq B(n - m) + (2c)^{-1} |\Sigma_{i=m+1}^{n} i^{\gamma} V_i| \quad .$$

Wegen $n - m \leq n \log m - m \log m \leq n \log n - m \log m$ ergibt sich somit aus (9):

$$(11) \qquad |A_m^{-1} - A_n^{-1}| \leq (2c)^{-1} |\Sigma_{i=m+1}^{n} i^{\gamma} V_i| + (B + 1)(n \log n - m \log m) \quad .$$

Nun können wir uns dem Beweis von (7) zuwenden:

Ist die Martingalbedingung (7.1) erfüllt, dann wollen wir Satz (6.1) anwenden. Dazu zerlegen wir die zu untersuchenden Partialsummen:

$$(12) \qquad \Sigma_{n=1}^{N} A_n U_n = \Sigma_{n=1}^{N} A_{n-1} U_n + \Sigma_{n=1}^{N} (A_n - A_{n-1}) U_n \quad ,$$

wobei wir hier und im folgenden $A_0 = 0$ vereinbaren. Da nach Konstruktion die Zufallsvariablen X_n A_n-meßbar sind (für $n \geq 2$ sind sie sogar A_{n-1}-meßbar), sind auch die D_n und damit die A_n A_n-meßbar. Weiterhin gilt nach Definition $|A_{n-1}| \leq (n-1)^{-1+\alpha} \leq 2n^{-1+\alpha}$ für $n \geq 2$, und da $\Sigma_{n=1}^{\infty} n^{-2+2\alpha}$ wegen $\alpha < \frac{1}{2}$ konvergiert, sind mit $W_n := A_{n-1}$ die Voraussetzungen des Satzes (6.1) erfüllt, und somit konvergiert $\Sigma_{n=1}^{\infty} A_{n-1} U_n$ f.s. Für die zweite Summe auf der rechten Seite der Formel (12) nutzen wir (8) und die Abschätzung (11) aus und erhalten für $n \geq 4$:

$$\begin{aligned} |A_n - A_{n-1}| &\leq A_n A_{n-1} ((2c)^{-1} |n^{\gamma} V_n| + (\beta + 1)(n \log n - (n-1)\log \\ &\leq C(n^{-2+2\alpha+\gamma} |V_n| + n^{-2+2\alpha} \log n) \quad , \end{aligned}$$

wobei die letzte Ungleichung mit einer geeigneten Konstante C (etwa $C = \max(c^{-1}, 4(\beta + 1)) < \infty$) wegen $A_{n-1} \leq 2n^{-1+\alpha}$ und $n \log n - (n-1)\log(n-1) \leq 1 + \log n \leq 2 \log n$ nach Mittelwertsatz gilt. Daraus folgt:

$$(13) \qquad \Sigma_{n=1}^{N} |(A_n - A_{n-1}) U_n| \leq C \Sigma_{n=1}^{N} n^{-2+2\alpha+\gamma} |U_n V_n| + C \Sigma_{n=1}^{N} n^{-2+2\alpha} \log n \quad ,$$

und für die zweite Summe auf der rechten Seite gilt:
$\Sigma_{n=1}^{N} n^{-2+2\alpha} \log n < \Sigma_{n=1}^{\infty} n^{-2+2\alpha} \log n < \infty$ wegen $\alpha < \frac{1}{2}$.
Weiterhin liefert der Satz von der monotonen Konvergenz (Beppo Levi):

$$\begin{aligned} E(\Sigma_{n=1}^{\infty} n^{-2+2\alpha+\gamma} |U_n V_n|) &= \Sigma_{n=1}^{\infty} n^{-2+2\alpha+\gamma} E(|U_n V_n|) \\ &\leq \sup_{n \in \mathbb{N}} E(|U_n V_n|) \; \Sigma_{n=1}^{\infty} n^{-2+2\alpha+\gamma} \quad . \end{aligned}$$

Wegen $\alpha \leq \frac{1}{4}$ (und nur hier geht im Martingalfall diese Einschränkung ein) konvergiert die letzte Reihe. Da $|U_nV_n| \leq U_n^2 + V_n^2$ gilt, folgt aus der Bedingung (7.1)

$$\sup_{n\in\mathbb{N}} E(|U_nV_n|) \leq \sup_{n\in\mathbb{N}} E(U_n^2) + \sup_{n\in\mathbb{N}} E(V_n^2) < \infty \quad ,$$

und insgesamt:

$$E(\Sigma_{n=1}^{\infty} n^{-2+2\alpha+\gamma}|U_nV_n|) < \infty \quad .$$

Ein wohlbekanntes Ergebnis ist nun (vgl. Loève (1977) I, S. 121):

$$\Sigma_{n=1}^{\infty} n^{-2+2\alpha+\gamma}|U_nV_n| < \infty \quad \text{f.s.}$$

Kehren wir nun zu (13) zurück, so erhalten wir:

$$\Sigma_{n=1}^{\infty} |(A_n - A_{n-1})U_n| < \infty \quad \text{f.s.} \quad ,$$

und die absolute Konvergenz zieht normale nach sich:

$$\Sigma_{n=1}^{\infty} (A_n - A_{n-1})U_n \quad \text{konvergiert f.s.}$$

Betrachten wir jetzt noch einmal (12), so erkennen wir, daß wir die Behauptung (7): $\Sigma_{n=1}^{\infty} A_nU_n$ konvergiert f.s. bewiesen haben.

Ist hingegen die Zuwachsbedingung (7.2) erfüllt, so wollen wir Satz (6.2) anwenden. Wir setzen hierzu $W_n := A_n$. Nach Definition gilt $A_n = O(n^{-1+\alpha})$, d.h. (6.2) (A) ist mit $\tau = \frac{1}{2} - \alpha > 0$ erfüllt. Für alle $\mu \geq 1$, alle k und alle $n \in N(k,\mu)$ wird (11) zu

$$|A_{m(k,\mu)}^{-1} - A_n^{-1}| \leq (2c)^{-1}|\Sigma_{i=m(k,\mu)+1}^{n} i^{\gamma}V_i| + (\beta+1)(n \log n - k^{\mu}\log k^{\mu})$$

Nun gilt $n \leq (k+2)^{\mu}$ für $n \in N'(k,\mu)$, folglich

$$n \log n - k^{\mu}\log k^{\mu} \leq \mu((k+2)^{\mu}\log(k+2) - k^{\mu}\log k)$$

und der Ausdruck auf der rechten Seite ist von der Größenordnung $O(k^{\mu-1}\log k)$. Weiterhin folgt mit Lemma (6.4) aus (7.2):

$$\sup_{n\in N(k,\mu)} |\Sigma_{i=m(k,\mu)+1}^{n} i^{\gamma} V_i| = O(k^{\gamma\mu+\frac{1}{2}\mu-\frac{1}{2}+\psi}) \quad \text{f.s.} \quad ,$$

und an dieser Größenordnung ändert sich nichts, wenn das Supremum über $N'(k,\mu)$ gebildet wird. Mit der Gleichung (8) ergibt sich unter Beachtung von $A_n \leq n^{-1+\alpha}$:

$$\sup_{n\in N'(k,\mu)} |A_n - A_{m(k,\mu)}| = O(k^{-2\mu+2\alpha\mu}(k^{\gamma\mu+\frac{1}{2}\mu-\frac{1}{2}+\psi} + k^{\mu-1}\log k)) \tag{14}$$

Nun gilt aber wegen $\alpha \leq \frac{1}{4}$:

$$-2\mu + 2\alpha\mu + \gamma\mu + \tfrac{1}{2}\mu - \tfrac{1}{2} + \psi \leq -\tfrac{1}{2}\mu - \tfrac{1}{2} - (\tfrac{1}{2} - \gamma)(\mu - 1) - (\tfrac{1}{2}$$

und

$$-2\mu + 2\alpha\mu + \mu - 1 \leq -\frac{1}{2}\mu - 1 \quad ,$$

und (6.2) (B) ist mit $\rho = \frac{1}{2} - \gamma - \psi$ erfüllt und es gilt $\psi < \rho < \frac{1}{2}$. Dann sind aber die Voraussetzungen des Satzes (6.2) erfüllt, so daß sich die Behauptung (7) ergibt.

Schritt 2:

Wir betrachten wieder einmal Realisationen ω mit $\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} x_n = \theta$ Dann gibt es nach (C) ein N , so daß für alle $n \geq N$ die Funktion M auf $[x_n - c_n , x_n + c_n]$ differenzierbar ist. Nach dem Mittelwertsatz existie also ξ_n mit

$$(2c_n)^{-1}(M(x_n + c_n) - M(x_n - c_n)) = M'(\xi_n) \quad ,$$

und es gilt $\lim_{n\to\infty} \xi_n = \theta$. Nutzen wir nun die Stetigkeit von M' in θ aus, so erhalten wir

$$\lim_{n\to\infty}(2c_n)^{-1}(M(x_n + c_n) - M(x_n - c_n)) = M'(\theta)$$

und mit der als Satz von Cesàro bekannten einfachsten Version des Lemma von Toeplitz:

$$\lim_{n\to\infty} n^{-1} \Sigma_{i=1}^{n} (2c_i)^{-1} (M(x_i + c_i) - M(x_i - c_i)) = M'(\theta) \quad .$$

Auf Grund von (1) gilt diese Aussage aber für fast alle ω . Weiterhin konvergiert $\Sigma_{n=1}^{\infty} n^{-1+\gamma} V_n$ nach Satz (6.1) bzw. Korollar (6.3) f.s., und mit dem Lemma von Kronecker folgt

$$n^{-1} \Sigma_{i=1}^{n} i^{\gamma} V_i \to 0 \quad \text{f.s.}$$

Damit ist aber bereits Behauptung (2) bewiesen:

$$D_n = n^{-1} \Sigma_{i=1}^{n} (2c_i)^{-1} (M(X_i + c_i) - M(X_i - c_i)) + (2cn)^{-1} \Sigma_{i=1}^{n} i^{\gamma} V_i$$

$$\to M'(\theta) \quad \text{f.s.}$$

Schritt 3:

Falls $M'(\theta) > 0$ gilt, ist eine einfache Folgerung aus (2):

(15) $\qquad nA_n \to M'(\theta)^{-1} \quad$ f.s.

Wir setzen $f := M$ und $E_n' := U_n + B_n$ mit $B_n = \frac{1}{2}(M(X_n + c_n) + M(X_n - c_n)) - M(X_n)$, so daß der Venter-Prozeß die Gestalt (3.1) besitzt. Nach Bemerkung ii) zu Satz (4.1) folgt aus (D) die Voraussetzung (4.1) (A) mit $\eta = 1$.

Um die systematischen Fehler B_n zu untersuchen, betrachten wir wieder Realisationen ω mit $\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} x_n = \theta$. Dann gibt es nach (D) ein N , so daß M'' für alle $n \geq N$ auf $[x_n - c_n , x_n + c_n]$ existiert und durch eine von n unabhängige Konstante C beschränkt ist. Nun können wir nach Taylor entwickeln:

$$M(x_n \pm c_n) = M(x_n) \pm c_n M'(x_n) + \frac{1}{2} c_n^2 M''(\xi_n^{\pm})$$

für gewisse ξ_n^+ und ξ_n^- mit $x_n - c_n < \xi_n^- < x_n < \xi_n^+ < x_n + c_n$, und erhalten die Abschätzung:

$$|b_n| = \frac{1}{4} c_n^2 |M''(\xi_n^+) + M''(\xi_n^-)| \leq \frac{1}{2} C c_n^2 \quad .$$

Da nach (1) $\lim_{n\to\infty} X_n(\omega) = \theta$ für fast alle ω gilt, ergibt sich somit

(16) $\qquad B_n = O(n^{-2\gamma})$ f.s.

Wegen $\gamma > \frac{1}{4}$ und $A_n = O(n^{-1})$ f.s. folgt daraus bereits

(17) $\qquad \Sigma_{n=1}^{\infty} n^{\delta} A_n B_n$ konvergiert f.s. für alle $\delta \leq \frac{1}{2}$,

wobei die Aussage für $\delta = \frac{1}{2}$ im fünften Schritt benötigt wird.

Gelingt es uns jetzt,

(18) $\qquad \Sigma_{n=1}^{\infty} n^{\delta} A_n U_n$ konvergiert f.s. für alle $\delta < \frac{1}{2}$

zu zeigen, dann können wir zwar noch nicht Satz (4.1) anwenden, da uns noch Voraussetzung (4.1) (C)

(19) $\qquad nA_n = M'(\theta)^{-1} + O(n^{-\tau})$ f.s. für ein $\tau > 0$

fehlt. Wir können jedoch Satz (3.2) benutzen, für den $nA_n \to M'(\theta)^{-1}$ f.s. genügt, und erhalten (5):

$$X_n = \theta + o(n^{-\delta}) \quad \text{f.s. für alle } \delta < \frac{1}{2} \quad .$$

Ist die Martingalbedingung (7.1) erfüllt, so zerlegen wir wie in (12)

$$\Sigma_{n=1}^{N} n^{\delta} A_n U_n = \Sigma_{n=1}^{N} n^{\delta} A_{n-1} U_n + \Sigma_{n=1}^{N} n^{\delta} (A_n - A_{n-1}) U_n \quad .$$

Nun sind die A_n immer noch $\mathcal{A}_n$-meßbar, aber wir besitzen mit (15) eine bessere Abschätzung:

$$n^{\delta} A_{n-1} = O(n^{\delta-1}) \quad \text{f.s.}$$

Da $\Sigma_{n=1}^{\infty} n^{-2+2\delta} < \infty$ für $\delta < \frac{1}{2}$ gilt, liefert uns Satz (6.1)

$$\Sigma_{n=1}^{\infty} n^{\delta} A_{n-1} U_n \quad \text{konvergiert f.s. für alle } \delta < \frac{1}{2} \quad .$$

Ebenso verbessert sich unsere Abschätzung für $|A_n - A_{n-1}|$ und wir erhalten analog zu (13)

$$\Sigma_{n=1}^{N} |n^{\delta}(A_n - A_{n-1})U_n| \leq C\Sigma_{n=1}^{N} A_n A_{n-1} n^{\delta+\gamma} |U_n V_n| + R$$

mit $R \leq C\Sigma_{n=1}^{\infty} n^{\delta} A_n A_{n-1} \log n < \infty$ für $\delta < 1$ wegen $n^{\delta} A_n A_{n-1} = O(n^{\delta-2})$ f.s. Mit denselben Argumentationen wie im ersten Schritt konvergiert $\Sigma_{n=1}^{\infty} n^{-2+\delta+\gamma} |U_n V_n|$ f.s. wegen $\delta + \gamma < 1$. Mit pfadweiser Anwendung des Kriteriums der majorisierten Konvergenz von Reihen erhalten wir somit

(20) $\quad \Sigma_{n=1}^{\infty} n^{\delta}(A_n - A_{n-1})U_n$ konvergiert f.s. für alle $\delta \leq \frac{1}{2}$,

und die Behauptung (18) folgt. Dabei haben wir (20) auch für $\delta = \frac{1}{2}$ bewiesen, da wir diese Aussage noch im fünften Schritt benötigen.

Ist die Zuwachsbedingung (7.2) erfüllt, dann setzen wir $W_n := n^{\delta} A_n$. Auf Grund von (15) ist (6.2) (A) mit $\tau = \frac{1}{2} - \delta > 0$ erfüllt. Wir zerlegen jetzt die Schwankungen in den Gewichten:

$$n^{\delta} A_n - m^{\delta} A_m = m^{\delta}(A_n - A_m) + (n^{\delta} - m^{\delta}) A_m$$

Setzen wir nun unsere verbesserte Abschätzung (15) in die Herleitung von (14) ein, so erhalten wir:

$$\sup_{n \in N'(k,\mu)} |A_n - A_{m(k,\mu)}| = O(k^{-\frac{3}{2}\mu + \gamma\mu - \frac{1}{2} + \psi} + k^{-\mu-1} \log k) \text{ f.s.}$$

Folglich gilt für alle $\delta \leq \frac{1}{2}$

$$\sup_{n \in N'(k,\mu)} m(k,\mu)^{\delta} |A_n - A_{m(k,\mu)}| = O(k^{-\frac{1}{2}\mu - \frac{1}{2} - \rho}) \quad \text{f.s.}$$

mit $\rho = \frac{1}{2} - \gamma - \psi < \frac{1}{2}$. Weiterhin gilt

$$\sup_{n \in N'(k,\mu)} (n^{\delta} - m(k,\mu)^{\delta}) |A_{m(k,\mu)}| = O(k^{\delta\mu - 1 - \mu}) \quad \text{f.s.} \ ,$$

so daß Voraussetzung (6.2) (B):

(21) $$\sup_{n\in N'(k,\mu)} |n^{\delta}A_n - m(k,\mu)^{\delta}A_{m(k,\mu)}| = O(k^{-\frac{1}{2}\mu-\frac{1}{2}-\rho}) \quad \text{f.s.}$$

mit $\rho = \frac{1}{2} - \gamma - \psi > \psi$ für alle $\delta \leq \frac{1}{2}$ erfüllt ist. Mit Satz (6.2) folgt dann die Behauptung (18). Auch hier haben wir die Aussage (21) zusätzlich für $\delta = \frac{1}{2}$ im Hinblick auf den fünften Schritt bewiesen.

Schritt 4:

Noch einmal betrachten wir Realisationen ω , und zwar diesmal mit $X_n(\omega) - \theta = x_n - \theta = O(n^{-\gamma})$. Dann gibt es nach (D) wieder ein N , so daß $x_n + c_n$ und $x_n - c_n$ für alle $n \geq N$ in einem offenen Intervall um θ liegen, auf dem M zweimal differenzierbar und M" beschränkt ist. Wir entwickeln nach Taylor, und zwar diesmal um θ:

$$M(x_n \pm c_n) = M(\theta) + (x_n \pm c_n)M'(\theta) + \frac{1}{2}(x_n \pm c_n)^2 M''(\eta_n^{\pm})$$

für gewisse η_n^+ und η_n^- aus dem offenen Intervall. Wir erhalten damit wegen $x_n \pm c_n = O(n^{-\gamma})$:

$$(2c_n)^{-1}(M(x_n + c_n) - M(x_n - c_n)) = M'(\theta) + O(n^{-\gamma}) \quad .$$

Da diese Aussage nach (5) für fast alle ω gilt, ergibt sich insgesamt

(22) $$n^{-1}\sum_{i=1}^{n}(2c_i)^{-1}(M(X_i + c_i) - M(X_i - c_i)) = M'(\theta) + O(n^{-\gamma}) \quad \text{f.}$$

Einsetzen in die Definition von D_n liefert wegen $\frac{1}{2} - 2\gamma < 0$ die Behauptung (4).

Schritt 5:

Wegen $M'(\theta) > 0$ ist es eine einfache Folgerung aus (2):

$D_n \to M'(\theta)$ f.s. und der Definition von A_n , daß für fast alle ω ein $N(\omega)$ existiert, so daß $A_n(\omega) = (nD_n(\omega))^{-1}$ für alle $n \geq N(\omega)$ gilt.

Mit dieser Überlegung ergibt sich sofort mit Hilfe von (6)

$$(23) \qquad nA_n - M'(\theta)^{-1} = nA_n M'(\theta)^{-1}(M'(\theta) - (nA_n)^{-1}) = O(n^{-\tau}) \quad \text{f.s.}$$

für alle $\tau < \frac{1}{2} - \gamma$. Damit ist die noch fehlende Voraussetzung (4.1) (C) erfüllt (vgl. (19)), und wir können Satz (4.1) anwenden:

$$n^{\frac{1}{2}}(X_{n+1} - \theta) = - n^{-\frac{1}{2}} \Sigma_{i=1}^{n} iA_i U_i - n^{-\frac{1}{2}} \Sigma_{i=1}^{n} iA_i B_i + o(1) \quad \text{f.s.}$$

Nun folgt mit dem Lemma von Kronecker aus Formel (17) mit $\delta = \frac{1}{2}$:

$$n^{-\frac{1}{2}} \Sigma_{i=1}^{n} iA_i B_i \to 0 \quad \text{f.s.}$$

Es bleibt also nur noch

$$(24) \qquad \Sigma_{n=1}^{\infty} n^{-\frac{1}{2}} (nA_n - M'(\theta)^{-1}) U_n \quad \text{konvergiert f.s.}$$

zu zeigen, und mit letztmaliger Anwendung des Lemmas von Kronecker folgt

$$n^{-\frac{1}{2}} \Sigma_{i=1}^{n} (iA_i - M'(\theta)^{-1}) U_i \to 0 \quad \text{f.s.}$$

und damit die Behauptung (3) (vgl. auch Korollar (4.2)).

Ist die Martingalbedingung (7.1) erfüllt, so zerlegen wir wieder:

$$\Sigma_{n=1}^{N} (n^{\frac{1}{2}} A_n - n^{-\frac{1}{2}} M'(\theta)^{-1}) U_n = \Sigma_{n=1}^{N} (n^{\frac{1}{2}} A_{n-1} - n^{-\frac{1}{2}} M'(\theta)^{-1}) U_n$$

$$+ \Sigma_{n=1}^{N} n^{\frac{1}{2}} (A_n - A_{n-1}) U_n \quad .$$

Nun ist $W_n := n^{\frac{1}{2}} A_{n-1} - n^{-\frac{1}{2}} M'(\theta)^{-1}$ A_{n-1}-meßbar und es gilt nach (15) und (23):

$$W_n = n^{-\frac{1}{2}} ((n-1)A_{n-1} - M'(\theta)^{-1}) + n^{-\frac{1}{2}} A_{n-1} = O(n^{-\frac{1}{2} - \tau}) \quad \text{f.s.}$$

für alle $\tau < \frac{1}{2} - \gamma$. Mit Satz (6.1) folgt dann die f.s. Konvergenz von $\Sigma_{n=1}^{\infty} W_n U_n$ und mit (20) für $\delta = \frac{1}{2}$ die Behauptung (24).

Ist die Zuwachsbedingung (7.2) erfüllt, dann setzen wir $W_n := n^{-\frac{1}{2}}(nA_n - M'(\theta)^{-1})$ und auf Grund von (23) ist (6.2) (A) für alle $\tau < \frac{1}{2} - \gamma$ erfüllt. Weiterhin gilt

$$W_n - W_m = n^{\frac{1}{2}}A_n - m^{\frac{1}{2}}A_m - (n^{-\frac{1}{2}} - m^{-\frac{1}{2}})M'(\theta)^{-1} \quad .$$

Wegen $\sup_{n\in N'(k,\mu)} |n^{-\frac{1}{2}} - m(k,\mu)^{-\frac{1}{2}}| = O(k^{-\frac{1}{2}\mu-1})$ folgt mit Hilfe der Formel (21) für $\delta = \frac{1}{2}$ die Voraussetzung (6.2) (B) mit $\rho = \frac{1}{2} - \gamma - \psi > \psi$ und der Satz (6.2) liefert uns die Behauptung (24). □

Wie bereits angekündigt, werden wir uns nicht sofort von diesem Beweis lösen, sondern wir werden untersuchen, welche Ergebnisse wir für $\gamma \leq \frac{1}{4}$ erhalten:

Im dritten Schritt wird Aussage (7.3) (17) genauer zu

$$\Sigma_{n=1}^{\infty} n^{\delta} A_n B_n \text{ konvergiert f.s. für alle } \delta < 2\gamma \quad ,$$

und an Stelle von (7.3) (5) ergibt sich

$$n^{\delta}(X_n - \theta) \to 0 \text{ f.s. für alle } \delta < 2\gamma \quad .$$

Setzen wir nun zusätzlich noch Stetigkeit von M" in θ voraus, so erhalten wir in (7.3) (22) unter Anwendung des Lemmas von Toeplitz einen Restterm der Größenordnung $o(n^{-\gamma})$ und damit an Stelle von (7.3) (4):

$$n^{\gamma}(D_n - M'(\theta)) = (2c)^{-1} n^{-1+\gamma} \Sigma_{i=1}^{n} i^{\gamma} V_i + o(1) \text{ f.s.}$$

Für $\gamma < \frac{1}{4}$ konvergiert $\Sigma_{n=1}^{\infty} n^{-1+2\gamma} V_n$ nach Satz (6.1) bzw. Korollar (6.3) f.s., so daß mit dem Lemma von Kronecker

$$n^{\gamma}(D_n - M'(\theta)) \to 0 \text{ f.s.}$$

für $\gamma < \frac{1}{4}$ folgt. Für $\gamma = \frac{1}{4}$ ergibt sich gerade Aussage (7.3) (4):

$$n^{\frac{1}{4}}(D_n - M'(\theta)) = (2c)^{-1} n^{-\frac{3}{4}} \Sigma_{i=1}^{n} i^{\frac{1}{4}} V_i + o(1) \quad f.s.$$

Für den fünften Schritt liefert uns nun Satz (4.1)

$$n^{2\gamma}(X_{n+1} - \theta) = - n^{-1+2\gamma} \Sigma_{i=1}^{n} iA_i U_i - n^{-1+2\gamma} \Sigma_{i=1}^{n} iA_i B_i + o(1) \quad f.s.$$

Da $nA_n - M'(\theta)^{-1} = O(n^{-\tau})$ f.s. für alle $\tau < \gamma$ und $B_n = O(n^{-2\gamma})$ f.s. gilt, konvergiert $\Sigma_{n=1}^{\infty} n^{-1+2\gamma}(nA_n - M'(\theta)^{-1})B_n$ f.s., und mit dem Lemma von Kronecker folgt:

$$n^{-1+2\gamma} \Sigma_{i=1}^{n} iA_i B_i = M'(\theta)^{-1} n^{-1+2\gamma} \Sigma_{i=1}^{n} B_i + o(1) \quad f.s.$$

Nutzen wir noch einmal die Stetigkeit von M" in θ aus, so erhalten wir für (7.3) (16) genauer:

$$n^{2\gamma} B_n \to \frac{1}{2} c^2 M''(\theta) \quad f.s.$$

und mit dem Lemma von Toeplitz:

$$n^{-1+2\gamma} \Sigma_{i=1}^{n} B_i \to (1 - 2\gamma)^{-1} \frac{1}{2} c^2 M''(\theta) \quad f.s.$$

Für $\gamma < \frac{1}{4}$ konvergiert $\Sigma_{n=1}^{\infty} n^{2\gamma} A_n U_n$ f.s. nach (7.3) (18), und mit dem Lemma von Kronecker ergibt sich insgesamt:

$$n^{2\gamma}(X_{n+1} - \theta) = - (2M'(\theta)(1 - 2\gamma))^{-1} c^2 M''(\theta) + o(1) \quad f.s.$$

für alle $\gamma < \frac{1}{4}$. Für $\gamma = \frac{1}{4}$ erhalten wir unter Verwendung von (7.3) (24):

$$n^{\frac{1}{2}}(X_{n+1} - \theta) = - M'(\theta)^{-1} n^{-\frac{1}{2}} \Sigma_{i=1}^{n} U_i - M'(\theta)^{-1} c^2 M''(\theta) + o(1) \quad f.s.$$

Gilt nun $M''(\theta) \neq 0$, so ergibt sich für $\gamma < \frac{1}{4}$ eine schlechtere Konvergenzrate von X_n als für $\gamma > \frac{1}{4}$, und für $\gamma = \frac{1}{4}$ ist die Konvergenzrate zwar genauso gut, jedoch tritt eine systematische Verfälschung der Gestalt $-M'(\theta)^{-1} c^2 M''(\theta)$ auf. Wie beim Kiefer-Wolfowitz-Prozeß (vgl. Satz (5.6)) können wir im Fall $M''(\theta) = 0$ auch kleinere Werte für γ zulassen, ohne daß eine systematische Verfälschung das asymptotische Verhalten der Störungen überlagert: Ist etwa M viermal differenzierbar in einer Um-

gebung von θ und $M^{(4)}$ stetig in θ , dann kann man zeigen, daß eine systematische Verfälschung der Gestalt $- (24\, M'(\theta)(1 - 4\gamma))^{-1} c^4 M^{(4)}(\theta)$ nur für $\gamma \leq \frac{1}{8}$ auftritt und für $\gamma > \frac{1}{8}$ die Repräsentation (7.3) (3) gilt.

Ein Vergleich mit der Arbeit von Venter (1967) zeigt, daß wir für eine sehr viel größere Klasse von zufälligen Störungen dieselben Konvergenzaussagen gewonnen haben. Der Vollständigkeit halber formulieren wir noch eine im höchsten Maße triviale Folgerung das asymptotische Verhalten betreffend:

(7.4) Korollar

Sind die Voraussetzungen des Satzes (7.3) erfüllt und die Störungen U_n genügen zusätzlich einem fast sicheren Invarianzprinzip:

(A) es ex. $\sigma > 0$ und eine Brown'sche Bewegung B , so daß gilt:

$$\Sigma_{i \leq t} U_i = \frac{1}{\sqrt{2}} \sigma B(t) + o(t^{\frac{1}{2}}) \quad \text{f.s.} \quad ,$$

dann gilt für den Venter-Prozeß

a) ein starkes Invarianzprinzip:

$$[t]^{\frac{1}{2}}(X_{[t]} - \theta) = (\sqrt{2}\, M'(\theta))^{-1} \sigma t^{-\frac{1}{2}} B(t) + o(1) \quad \text{f.s.} \quad ;$$

b) ein Gesetz vom iterierten Logarithmus:

$$\limsup_{n \to \infty} \frac{n^{\frac{1}{2}} |X_n - \theta|}{\sqrt{2 \log\log n}} = (\sqrt{2}\, M'(\theta))^{-1} \sigma \quad \text{f.s.} \quad ;$$

c) ein Verteilungsinvarianzprinzip:

$$(tn^{\frac{1}{2}}(X_{[nt]} - \theta))_{t \in [0,1]} \xrightarrow{\mathcal{D}} (\sqrt{2}\, M'(\theta))^{-1} \sigma B \quad (\text{in } \mathcal{D}[0,1]) \quad ;$$

d) asymptotische Normalität:

$$n^{\frac{1}{2}}(X_n - \theta) \xrightarrow{\mathcal{D}} Y$$

mit $Y \sim N(0, \frac{1}{2} M'(\theta)^{-2}\sigma^2)$.

Diese Aussagen folgen sofort aus (7.3) (3) mit den Ergebnissen des Anhangs über Invarianzprinzipien. Der Faktor $\frac{1}{\sqrt{2}}$ im fast sicheren Invarianzprinzip (7.4) (A) ist wieder eingeführt worden, um Übereinstimmung mit dem Fall der bedingten Unabhängigkeit zu erzielen, da $U_n = \frac{1}{2}(U_n^+ + U_n^-)$ gilt (vgl. Kapitel 1 und die Bemerkung zum Korollar (5.5)). Beachten wir nun noch, daß zu jedem Zeitpunkt zwei Beobachtungen durchgeführt werden, so müssen wir das asymptotische Verhalten von $(2n)^{\frac{1}{2}}(X_n - \theta)$ untersuchen, um mit dem Robbins-Monro-Prozeß vergleichen zu können. Nach (7.4) d) ist aber $(2n)^{\frac{1}{2}}(X_n - \theta)$ asymptotisch normalverteilt mit Erwartungswert 0 und Varianz $M'(\theta)^{-2}\sigma^2$, und dieses ist gerade die beim Robbins-Monro-Prozeß mögliche minimale Varianz bei optimaler Wahl von $a = M'(\theta)^{-1}$.

Zum Abschluß sei noch darauf hingewiesen, daß für gewisse schwach abhängige Folgen von Störungen, die die Momentenbedingung (6.5) (A) (vgl. Anhang III) und damit die Zuwachsbedingung (7.2) erfüllen, zumindest Verteilungsinvarianzprinzipien und asymptotische Normalität gelten. Neuere Ergebnisse aus diesem Problemkreis wurden z.B. von Peligrad (1982) und Herrndorf (1984) für ρ-mischende (vollständig reguläre) Folgen bewiesen.

Kapitel 8: Vermischtes

Zum Abschluß werden wir noch einige vielleicht interessante Fragestellungen ansprechen:

Mehrdimensionale stochastische Approximation:

Eine naheliegende Verallgemeinerung der Problemstellung ist es, die Nullstelle θ einer mehrdimensionalen Funktion $f: \mathbb{R}^N \to \mathbb{R}^N$ bzw. das Maximum oder Minimum einer reellen Funktion mehrerer Veränderlicher $M: \mathbb{R}^N \to \mathbb{R}$ zu bestimmen. Dabei kann das Problem des Auffindens eines Extremums für differenzierbares M wieder auf das Bestimmen einer Nullstelle der Ableitung von M , welche bekanntlich eine Funktion von $\mathbb{R}^N$ nach $\mathbb{R}^N$ ist, zurückgeführt werden.
In Analogie zum eindimensionalen Fall kann man auf dem Fixpunktprinzip basierende Algorithmen der Gestalt

$$X_{n+1} = X_n - a_n Y_n$$

mit

$$Y_n = f(X_n) + E_n + B_n$$

betrachten, wobei die X_n , die Beobachtungen Y_n , die zufälligen Störungen E_n und die systematischen Fehler B_n jetzt zufällige N-dimensionale Vektoren und die Schrittweiten a_n reelle Zahlen sind. Ein N-dimensionaler Robbins-Monro-Prozeß:

$$Y_n = f(X_n) + U_n$$

und ein N-dimensionaler Kiefer-Wolfowitz-Prozeß:

$$Y_n = c_n^{-1}(f_n(X_n) + U_n)$$

mit

$$f_n(x) = \frac{1}{2}(M(x + c_n\varepsilon_1) - M(x - c_n\varepsilon_1),\ldots,$$

$$M(x + c_n\varepsilon_N) - M(x - c_n\varepsilon_N))^T$$

oder

$$f_n(x) = (M(x + c_n\varepsilon_1) - M(x),\ldots,M(x + c_n\varepsilon_N) - M(x))^T \quad ,$$

wobei $\varepsilon_1,\ldots,\varepsilon_N$ orthonormale Basisvektoren des $\mathbb{R}^N$ sind und T das Transpositionszeichen ist, wurde erstmals von Blum (1954 b) untersucht, der unter geeigneten Voraussetzungen fast sichere Konvergenz nachweisen konnte. Im wesentlichen muß dabei für die Funktion f vorausgesetzt werden, daß sie die Ableitung einer differenzierbaren Funktion von $\mathbb{R}^N$ nach $\mathbb{R}$ ist, oder daß zumindest eine Lyapunov-Funktion $V: \mathbb{R}^N \to \mathbb{R}$, d.h. es gilt $(\frac{d}{dx} V(x))^T f(x) < 0$ für alle $x \neq \theta$, existiert. Im weiteren Verlauf konnten fast alle Ergebnisse für eindimensionale stochastische Approximationsverfahren auch auf mehrdimensionale übertragen werden, und $\mathbb{R}^N$ konnte sogar durch einen beliebigen separablen Hilbertraum ersetzt werden (vgl. z.B. Walk (1977)).
Ljung (1978) untersuchte dann erstmals Algorithmen der Gestalt

$$(8.1) \qquad X_{n+1} = X_n - A_n(f(X_n) + E_n + B_n) \quad ,$$

wobei die Schrittweiten A_n (reellwertige) Zufallsvariablen sein dürfen, und verzichtete auf die Forderung einer Martingalbedingung für die Störungen E_n . Wir werden sein Ergebnis in etwas vereinfachter Form zitieren:

(8.2) Satz: (Ljung (1978))

Genügen die Funktionen $f: \mathbb{R}^N \to \mathbb{R}^N$, $V: \mathbb{R}^N \to \mathbb{R}$, $\theta \in \mathbb{R}^N$, die Zufallsvariablen A_n mit $A_n > 0$ und die zufälligen N-dimensionalen

Vektoren E_n und B_n den Voraussetzungen

(A) V ist zweimal stetig differenzierbar und $\frac{d}{dx} V(x) = - f(x)$,

(B) $\{x|V(x) \geq C\}$ ist kompakt für alle $C > \inf_x V(x)$,

(C) $f(\theta) = 0$ und $f(x) \neq 0$ für $x \neq \theta$,

(D) $\sup_x ||\frac{d}{dx} f(x)|| < \infty$,

(E) $\inf_{||x-\theta|| \geq \delta} ||f(x)|| > 0$ für alle $\delta > 0$,

(F) $A_n \to 0$ f.s. und $\Sigma_{n=1}^{\infty} A_n = \infty$ f.s. ,

(G) $B_n \to 0$ f.s. ,

(H) $\Sigma_{n=1}^{\infty} A_n E_n$ konvergiert f.s. ,

dann gilt für den durch (8.1) definierten Prozeß:

$$X_n \to \theta \quad \text{f.s.}$$

Unter der Annahme deterministischer Schrittweiten $A_n = n^{-1}$ formulierte Ruppert (1982) eine starke Repräsentation des allgemeinen N-dimensional Prozesses, die meines Erachtens sogar in folgender etwas leichter lesbaren Form gilt:

(8.3) Satz: (Ruppert (1982))

Genügen $f: \mathbb{R}^N \to \mathbb{R}^N$, $\theta \in \mathbb{R}^N$, eine N×N-Matrix D , die zufälligen N-dimensionalen Vektoren X_1, E_n und B_n , sowie der durch

$$X_{n+1} = X_n - n^{-1}(f(X_n) + E_n + B_n)$$

definierte Prozeß und Δ mit $0 < \Delta \leq \frac{1}{2}$ den Voraussetzungen:

(A) es ex. $\eta > 0$ mit $f(x) = D(x - \theta) + O(||x - \theta||^{1+\eta})$ für $x \to \theta$,

(B) $X_n \to \theta$ f.s. ,

(C) es ex. $\beta \in \mathbb{R}^N$ und $\rho > 0$ mit $n^{\Delta}B_n = \beta + O(n^{-\rho})$ f.s. ,

(D) $\Sigma_{n=1}^{\infty} n^{-M}E_n$ konvergiert f.s. für alle N×N-Matrizen M mit $\lambda_{min}(M) > 1 - \Delta$,

(E) $\lambda_{min}(D) > \Delta$,

dann gilt:

$$n^{\Delta}(X_{n+1} - \theta) = -(D - \Delta I)^{-1}\beta - n^{-D+\Delta}\Sigma_{i=1}^{n} i^{D-I}E_i + O(n^{-\varepsilon}) \quad \text{f.s.}$$

für ein $\varepsilon > 0$.

Dabei bezeichnet I die N×N-Einheitsmatrix , $\lambda_{min}(M)$ den kleinsten Realteil der Eigenwerte von M und $n^M = \exp((\log n)M)$ mit $\exp(M) = \Sigma_{i=0}^{\infty}(i!)^{-1}M^i$. Es stellt sich nun im Hinblick auf die Ergebnisse im eindimensionalen Fall die erste wahrscheinlich noch leicht zu beantwortende Frage:

(I) Läßt sich die Aussage des Satzes (8.3) auf Algorithmen der Gestalt (8.1), d.h. mit reellwertigen zufälligen Schrittweiten A_n , verallgemeinern ?

Da es recht willkürlich ist, für alle Dimensionen gleiche Schrittweiten zu fordern, erscheint es sinnvoll, für die "Schrittweiten" N×N-Matrizen zuzulassen. Wir betrachten also an Stelle von (8.1) Algorithmen der Form

(8.4) $$X_{n+1} = X_n - A_n(f(X_n) + E_n + B_n) \quad ,$$

wobei die A_n zufällige N×N-Matrizen sind. Nevel'son und Khas'minskii

(1973/74) behandelten eine derartige Verallgemeinerung des Robbins-Monro-Prozesses:

Sind die $A_n = n^{-1}A$ deterministisch mit fester N×N-Matrix A , dann ist unter gewissen zusätzlichen Voraussetzungen $n^{\frac{1}{2}}(X_n - \theta)$ asymptotisch normalverteilt mit Erwartungswert 0 und Kovarianzmatrix $\Sigma(A)$. Ist $D := (\frac{d}{dx} f(x)|_{x=\theta})$ nichtsingulär, dann ist $\Sigma(A)$ "minimal" im positiv-definiten Sinne, falls $A = D^{-1}$ gilt, d.h. $\Sigma(A) - \Sigma(D^{-1})$ ist nichtnegativ-definit. Entsprechend dem Venter-Prozeß läßt sich wieder ein adaptives Verfahren der Gestalt (8.4) konstruieren, so daß nA_n f.s. gegen D^{-1} konvergiert, und im Falle bedingt unabhängiger Störungen konnten Nevel'son und Khas'minskii (1973/74) asymptotische Normalität von $n^{\frac{1}{2}}(X_n - \theta)$ mit Kovarianzmatrix $(2N)^{-\frac{1}{2}}\Sigma(D^{-1})$ nachweisen, so daß gleiche Güte wie bei optimaler Wahl von $A = D^{-1}$ erzielt wird. Der Faktor $(2N)^{-\frac{1}{2}}$ erklärt sich wieder daraus, daß zu jedem Zeitpunkt 2N Beobachtungen gemacht werden. Es stellt sich die zweite Frage:

(II) Lassen sich die Aussagen der Sätze (8.2) und (8.3) auch auf Prozesse der Form (8.4) verallgemeinern ?

Ein Exkurs über notwendige und hinreichende Bedingungen

Für den (eindimensionalen) Robbins-Monro-Prozeß läßt sich unter einer zusätzlichen Regularitätsbedingung an die zugrunde liegende Funktion eine recht einfache Bedingung angeben, und zwar ist die f.s. Konvergenz des Robbins-Monro-Prozesses äquivalent zur Gültigkeit eines starken Gesetzes der großen Zahlen für die zufälligen Störungen:

(8.5) Satz:

Genügen $M: \mathbb{R} \to \mathbb{R}$ und $\theta \in \mathbb{R}$ den Voraussetzungen

(A) $M(x)(x - \theta) > 0$ für alle $x \neq \theta$,

(B) $\inf_{|x-\theta|>\delta} |M(x)| > 0$ für alle $\delta > 0$,

(C) M ist stetig differenzierbar mit $\sup_x |M'(x)| < \infty$,

dann gilt für die Zufallsvariablen U_n und den durch

$$X_{n+1} = X_n - an^{-1}(M(X_n) + U_n) \quad ,$$

$a > 0$, definierten Robbins-Monro-Prozeß:

$$X_n \to \theta \quad \text{f.s.}$$

genau dann, wenn

$$n^{-1} \Sigma_{i=1}^{n} U_i \to 0 \quad \text{f.s.}$$

Beweis:

Die eine Richtung ist fast trivial:

Gilt $X_n \to \theta$ f.s., dann iterieren wir die Rekursionsformel:

$$X_{n+1} = X_1 - a\Sigma_{i=1}^{n} i^{-1}(M(X_i) + U_i)$$

und erhalten somit:

$$\Sigma_{n=1}^{\infty} n^{-1}(M(X_n) + U_n) \quad \text{konvergiert f.s.}$$

Das Lemma von Kronecker liefert wieder:

$$n^{-1} \Sigma_{i=1}^{n}(M(X_i) + U_i) \to 0 \quad \text{f.s.}$$

Nun folgt aber aus $X_n \to \theta$ f.s. und der Stetigkeit von M , daß $M(X_n)$ f.s. gegen 0 konvergiert. Mit dem Satz von Cesàro gilt:

$$n^{-1} \Sigma_{i=1}^{n} M(X_i) \to 0 \quad \text{f.s.}$$

und folglich das starke Gesetz der großen Zahlen:

$$n^{-1} \Sigma_{i=1}^{n} U_i \to 0 \quad \text{f.s.}$$

Die Rückrichtung ist sehr viel komplizierter, und wir werden sie nur andeuten: Wir können ausnutzen, daß Ljung (1978) den Satz (8.2) unter der schwächeren Voraussetzung

(H') für die durch $Z_{n+1} = Z_n + A_n(E_n - Z_n)$, $Z_1 = 0$ definierten Zufallsvariablen gilt: $Z_n \to 0$ f.s.

an Stelle von (H) bewiesen hat. Es sei bemerkt, daß wir tatsächlich (H') aus (H) folgern können, indem wir $(Z_n)_{n\in\mathbb{N}}$ als stochastischen Approximationsprozeß zur Bestimmung der Nullstelle 0 von $f = id$ mit Störungen $-E_n$ identifizieren und Satz (2.2) anwenden. Im vorliegenden Fall des Robbins-Monro-Prozesses ergibt sich für die Z_n

$$Z_{n+1} = n^{-1} \Sigma_{i=1}^{n} U_i \quad ,$$

denn es gilt nach Voraussetzung $Z_1 = 0$ und induktiv

$$nZ_{n+1} = nZ_n + (U_n - Z_n) = (n-1)Z_n + U_n = \Sigma_{i=1}^{n} U_n \quad .$$

Das starke Gesetz der großen Zahlen ist also gerade (H') und somit auch hinreichend für $X_n \to \theta$ f.s., da die anderen Voraussetzungen des Satzes (8.2) erfüllt sind. □

Verzichten wir auf die Voraussetzung (8.5) (B), wie wir es in den vorangegangenen Kapiteln getan haben, dann ist das starke Gesetz der großen Zahlen nicht mehr hinreichend:

Ist M eine Funktion, die (8.5) (A) und (C) erfüllt, mit $\lim_{x\to\infty} M(x) = 0$,

etwa M mit $M(x) = xe^{-x^2}$, und ist $X_1 = U_1 = 0$ und $U_n = - M(X_n) - (\log n)^{-1}$, $n \geq$
Dann ergibt sich $X_{n+1} = X_n + a(n \log n)^{-1}$ und nach Iteration:
$X_{n+1} = a \sum_{i=2}^{n} (i \log i)^{-1}$. Folglich gilt $X_n \to \infty$. Nach Voraussetzung folgt $M(X_n) \to 0$ und damit auch $U_n \to 0$. Mit dem Satz von Cesâro gilt somit $n^{-1} \sum_{i=1}^{n} U_n \to 0$, aber X_n konvergiert nicht gegen $\theta = 0$.

Die Aussage des Satzes (8.5) läßt sich ohne Schwierigkeiten auf den N-dimensionalen Fall übertragen. Eine allgemeinere Behandlung für den Robbins-Monro-Prozeß findet man bei Clark (1984). Es stellt sich als dritte Frage:

(III) Gibt es auch für andere, insbesondere adaptive stochastische Approximationsprozesse sinnvolle hinreichende und notwendige Bedingungen ? Und sind diese in etwa von der Form (8.5) (H') ?

Rekursive Schätzung

Für die Folge $(X_n)_{n \in \mathbb{N}}$ der arithmetischen Mittel $X_{n+1} = n^{-1} \sum_{i=1}^{n} Z_i$ von Beobachtungen Z_n , $n \in \mathbb{N}$, ist bekannt, daß sie sich recht einfach rekursiv berechnen läßt:

$$X_{n+1} = X_n - n^{-1}(X_n - Z_n)$$

mit $X_1 = 0$. Wir können also die Folge der arithmetischen Mittel als Robbins-Monro-Prozeß zur Bestimmung der Nullstelle θ der bis auf diesen Lokationsparameter bekannten Funktion f mit $f(x) = x - \theta$ interpretieren, wobei bei der Beobachtung $Y_n := X_n - Z_n$ von f an der Stelle X_n zufällige Störungen $U_n := \theta - Z_n$ auftreten. Jedoch liefern unsere Ergebnisse natürlich keine neuen Erkenntnisse für das arithmetische Mittel.

Allgemein läßt sich rekursive Schätzung mit den Methoden der stochastischen Approximation immer dann durchführen, wenn der funktionale Zusammenhang zwischen den einstellbaren Kontrollgrößen X_n und den "fehlerfreien" Beobachtungen $Y_n - U_n = f(X_n)$ bis auf einige wenige Parameter bekannt ist, zu denen offenkundig auch das gesuchte θ mit $f(\theta) = 0$ gehören sollte.
Ist etwa f linear, d.h. $f(x) = \lambda(x - \theta)$ mit $\lambda > 0$, und sind die zufälligen Störungen $U_n = Y_n - f(X_n)$ unabhängig und identisch verteilt mit Erwartungswert 0 und Varianz σ^2 , dann wissen wir (vgl. Kapitel 5), daß die durch

$$X_{n+1} = X_n - A_n Y_n$$

mit $Y_n = \lambda(X_n - \theta) + U_n$ definierte rekursive Schätzung minimale Varianz in der Grenzverteilung von $n^{\frac{1}{2}}(X_n - \theta)$ für $A_n = \lambda^{-1}n^{-1}$ besitzt, falls λ bekannt ist. Falls λ unbekannt ist, können wir mit einem adaptiven Verfahren, für das $nA_n \to \lambda^{-1}$ f.s. gilt, dieselbe Güte erzielen. Dazu können wir z.B. das im Kapitel 7 beschriebene Venter-Verfahren benutzen. Sind weiterhin die U_n absolut stetig und ist die Dichte g von U_n bekannt, dann können wir durch eine Transformation h der Beobachtungen Y_n und Betrachtung des durch

$$X_{n+1} = X_n - A_n h(Y_n)$$

definierten Prozesses die Varianz der Grenzverteilung weiter verkleinern. Gilt $h = -g'g^{-1}$, dann ist diese Varianz minimal und nimmt den Wert $I(g)^{-1}\lambda^{-2}$ an, wobei I die Fisher-Information ist. Das bedeutet insbesondere, daß für diese Wahl von h die Varianz in der Grenzverteilung nicht nur minimal in der Klasse der stochastischen Approximationsverfahren, sondern auch in der Klasse aller erwartungstreuen [illegible]verfahren ist. Dabei sei bemerkt, daß die letzte Aussage

lediglich bedeutet, daß die Varianz der Grenzverteilung die Rao-Cramér-Schranke für den vorliegenden Fall annimmt. Ist g unbekannt, dann können wir diese Güte auch noch sichern, wenn wir das Verfahren noch in einem zweiten Sinne adaptiv abändern, indem wir h durch eine geeignete Folge zufälliger Funktionen H_n ersetzen:

$$X_{n+1} = X_n - A_n H_n(Y_n) \quad .$$

Dieses Ergebnis wurde von Fabian (1973) bewiesen.

Eine allgemeinere Behandlung rekursiver Schätzverfahren kann man bei Nevel'son und Has'minskii (1973) finden. Wir werden hier jedoch zum Abschluß ein Ergebnis von Holst (1984) über rekursive robuste Verfahren mit absolut regulären (ß-mischenden) Beobachtungen vorstellen und vervollständigen (zur Definition von ß-mischend vgl. Anhang III): Für eine Folge $(Z_n)_{n\in\mathbb{N}}$ unabhängiger, identisch verteilter Zufallsvariablen wurde von Huber (vgl. Huber (1981), S. 43 ff) als robuste Schätzfunktion für den Lokationsparameter θ der M-Schätzer T_n vorgeschlagen, welcher als Lösung der impliziten Gleichung

$$\Sigma_{i=1}^{n} \psi(Z_i - T_n) = 0$$

definiert ist, wobei ψ eine geeignet zu wählende Funktion ist (wir wollen auf Stärken und Schwächen dieser Definition hier nicht weiter eingehen). Ist ψ die identische Abbildung, dann ist T_n gerade das arithmetische Mittel $n^{-1} \Sigma_{i=1}^{n} Z_i$, welches bekanntlich nicht robust ist. Es liegt trotzdem nahe, auch für andere Funktionen ψ den M-Schätzer T_n zumindest näherungsweise durch eine Rekursionsformel zu berechnen:

$$T_n^* = T_{n-1}^* + A_n \psi(Z_n - T_{n-1}^*)$$

mit $T_0^* = 0$. Martin und Masreliez (1975) konnten nachweisen, daß unter geeigneter Wahl von A_n die rekursiv erzeugte Schätzfunktion T_n^* asymptotisch dieselbe Güte wie T_n besitzt. A_n geeignet bedeutet dabei, daß entweder $A_n = -n^{-1}\lambda^{-1}$ mit $\lambda = M'(0)$ gilt, wobei M durch $M(x) := E(\psi(Z_n - x))$ definiert sein soll, oder daß dieser Zusammenhang zumindest in der Form $nA_n \to -\lambda^{-1}$ f.s. gilt, was wiederum zu adaptiven Prozessen führt.

Für den Fall, daß die Beobachtungen Z_n lediglich stationär und ß-mischend sind, konnte Holst (1984) unter zusätzlichen Regularitätsbedingungen zumindest fast sichere Konvergenz von T_n^* gegen θ und Aussagen über die Konvergenzgeschwindigkeit für adaptive Verfahren beweisen. Mit Hilfe unseres Satzes (4.1) können wir diese Ergebnisse noch durch eine fast sichere Repräsentation abrunden. Der Übersichtlichkeit halber formulieren wir die Aussage unter vereinfachenden Einschränkungen an die Mischungskoeffizienten $(ß(k))_{k\in\mathbb{N}}$.

(8.6) Satz:

Genügen eine beschränkte Funktion $\psi: \mathbb{R} \to \mathbb{R}$, $\theta \in \mathbb{R}$, eine stationäre Folge ß-mischender Zufallsvariablen $(Z_n)_{n\in\mathbb{N}}$ mit Mischungskoeffizieten $(ß(k))_{k\in\mathbb{N}}$ sowie die durch

$$M(x) := E(\psi(Z_n - x))$$

definierte Funktion $M: \mathbb{R} \to \mathbb{R}$ und ein γ mit $0 < \gamma < \frac{1}{2}$ den Voraussetzungen:

(A) ψ ist nichtfallend ,

(B) $M(\theta) = 0$ und $M(x) \neq 0$ für $x \neq \theta$,

(C) M ist Lipschitz-stetig ,

(D) $ß(k) = O(k^{-\beta})$ für ein $\beta > \gamma(1-\gamma)^{-1}$,

dann gilt für den durch

$$T_n^* = T_{n-1}^* + A_n \psi(Z_n - T_{n-1}^*)$$

mit $T_0^* := 0$,

$$D_n = n^{-1} \sum_{i=1}^{n} (2c_i)^{-1} (\psi(Z_i - T_{i-1}^* + c_i) - \psi(Z_i - T_{i-1}^* - c_i)) \quad ,$$

$c_n = n^{-\gamma}$, und

$$A_n = \begin{cases} (n\mu)^{-1} & \text{für } D_n < \mu \\ (nD_n)^{-1} & \text{sonst} \end{cases} ,$$

$\mu > 0$, definierten Prozeß:

(1) $\quad T_n^* \to \theta$ f.s.

Ist zusätzlich

(E) M ist differenzierbar in einer Umgebung von θ ,

und M' ist stetig in θ ,

erfüllt, dann gilt:

(2) $\quad D_n \to M'(\theta)$ f.s.

Sind an Stelle von (D) und (E) die stärkeren Voraussetzungen

(D*) $\beta(k) = O(k^{-\beta})$ für alle $\beta > 0$,

(E*) M ist differenzierbar in einer Umgebung von θ ,

und es ex. $\eta > 0$ mit

$M(x) = M'(\theta)(x - \theta) + O(|x - \theta|^{1+\eta})$ für $x \to \theta$,

und zusätzlich

(F) $\quad - M'(\theta) > \mu$

erfüllt, dann gilt:

(3) $\quad n^{\delta}(T_n^* - \theta) \to 0$ f.s.

für alle $\delta < \frac{1}{2}$ und

$$(4) \qquad n^{\frac{1}{2}}(T_n^* - \theta) = n^{-\frac{1}{2}} \Sigma_{i=1}^{n} iA_i(\psi(Z_i - T_{i-1}^*) - M(T_{i-1}^*)) + O(n^{-\varepsilon}) \quad \text{f.s.}$$

für ein $\varepsilon > 0$.

Bemerkungen:

i) Die Aussagen (1) bis (3) wurden von Holst (1984) mit ähnlichen Methoden wie den von uns in Kapitel 2 und 3 vorgestellten bewiesen, wobei jedoch die spezielle Mischungseigenschaft im Vordergrund stand. Eine Aussage in der Art von (4) wurde von Holst (1984) bereits in Kenntnis von Ruppert's (1982) Arbeit erwartet; mit Hilfe unseres Satzes (4.1) ist es nun möglich, diese zu beweisen.

ii) Die Beschränktheit von ψ hat drei angenehme Effekte: Erstens ist M wohldefiniert, zweitens sind die auftretenden zufälligen Störungen $U_n := \psi(Z_n - T_{n-1}^*) - M(T_{n-1}^*)$ gleichmäßig beschränkt, und drittens wächst D_n höchstens wie n^{γ} . Weiterhin ist auf Grund von (A) die Funktion M nichtwachsend, welches zusammen mit (B) die Eigenschaft $M(x)(x - \theta) < 0$ für $x \neq \theta$ liefert.

iii) Die Argumentation bei Holst (1984) können wir insofern verschärfen, daß wir statt (D*) , was z.B. erfüllt ist, wenn $\beta(k)$ exponentiell fällt, lediglich $\beta(k) = O(k^{-\beta})$ für ein $\beta \geq (1 + 2\gamma)(1 - 2\gamma)^{-1}$ fordern müssen. Das bedeutet insbesondere, daß wir durch Verkleinerung von γ größere Abhängigkeiten zulassen können. Da natürlich uns unsere Funktion ψ bekannt ist, treten anders als beim Venter-Prozeß (vgl. Kapitel 7) keine systematischen Fehler auf, so daß wir lediglich als untere Schranke $\gamma \geq 0$ voraussetzen müssen. Vergleich mußte beim

Venter-Prozeß $\gamma > \frac{1}{4}$ gelten, um eine optimale Konvergenzrate zu erzielen.

iv) Ist ψ differenzierbar und ψ' beschränkt und Lipschitz-stetig, dann kann D_n durch

$$D_n' = n^{-1} \Sigma_{i=1}^{n} \psi'(Z_i - T_{i-1}^*)$$

ersetzt werden. Da keine c_n auftreten, genügt es nach Holst (1984), an Stelle von (D) die schwächere Forderung

(D') $\beta(k) = O((\log k)^{-1-\beta})$ für ein $\beta > 0$

und an Stelle von (D*)

(D") $\beta(k) = O(k^{-1})$

vorauszusetzen.

Beweis:

Wie bereits bemerkt, wurden die Aussagen (1) bis (3) von Holst (1984) bewiesen. Wir werden diesen Beweis hier nicht nachvollziehen, sondern vielmehr sogar beim Nachweis von (4) einige Lemmas lediglich erwähnen. Setzen wir $f := -M$, $X_n := T_{n-1}^*$ und $E_n' := -(\psi(Z_n - X_n) - M(X_n))$, dann besitzt der betrachtete Prozeß die Gestalt (3.1). Nach den Lemmas 3.2, 3.5 und 3.6 von Holst (1984) gilt

$$\Sigma_{n=1}^{\infty} n^{\delta} A_n E_n' \text{ konvergiert f.s. für alle } \delta < \frac{1}{2} .$$

Weiterhin ergibt sich aus (3) und (E*)

$$(2c_n)^{-1}(M(X_n - c_n) - M(X_n + c_n)) + M'(\theta) = O(n^{-\gamma\eta}) \quad \text{f.s.}$$

und mit dem Lemma von Kronecker:

$$n^{-1} \Sigma_{i=1}^{n} (2c_i)^{-1}(M(X_i - c_i) - M(X_i + c_i)) = -M'(\theta) + O(n^{-\rho'}) \quad \text{f.s.}$$

für ein $\rho' > 0$. Unter Ausnutzung der Abschätzungen in Lemma 4.1 von Holst (1984) folgt für ein $\rho > 0$:

$$D_n = - M'(\theta) + O(n^{-\rho}) \quad \text{f.s.}$$

und über den bekannten Zwischenschritt

$$D_n^{-1} + M'(\theta)^{-1} = (D_n M'(\theta))^{-1}(D_n + M'(\theta)) = O(n^{-\rho}) \quad \text{f.s.}$$

die Voraussetzung (4.1) (C):

$$nA_n = - M'(\theta)^{-1} + O(n^{-\rho}) \quad \text{f.s.}$$

Somit können wir Satz (4.1) mit $\Delta = \frac{1}{2}$ anwenden und erhalten Behauptung (4). □

Wünschenswert wäre gewesen, an Stelle von (8.6) (4) das Ergebnis

$$(8.7) \qquad n^{\frac{1}{2}}(T_n^* - \theta) = - M'(\theta)^{-1} n^{-\frac{1}{2}} \Sigma_{i=1}^{n} (\psi(Z_i - T_{i-1}^*) - M(T_{i-1}^*)) + o(1)$$

zu erreichen. Es stellt sich somit die Frage:

(IV) Unter welchen eventuell zusätzlichen Voraussetzungen gilt (8.7) ? Und wann folgt daraus asymptotische Normalität oder sogar mehr ?

Eine weitere und für diese Arbeit letzte Frage in diesem Zusammenhang ist:

(V) Kann auf die Vorkenntnis (8.6) (F) über die Größe von $M'(\theta)$ verzichtet werden, wenn A_n anders gestutzt wird (vgl. Venter-Prozeß) ?

Anhang

I) Hilfsmittel aus der Analysis:

In diesem Abschnitt werden wir drei kleine Lemmas angeben. Das erste beinhaltet eine in der vorliegenden Arbeit häufig benutzte Integralabschätzung für Summen, das zweite befaßt sich mit einer im zweiten Teil des Anhangs benötigten Abschätzung für das Supremum einer lokal beschränkten Funktion, und mit dem dritten werden wir eine unbedeutende Lücke in Ruppert's (1982) Beweis zur Konvergenzgeschwindigkeit schließen (vgl. Bemerkung i) zu Satz (3.2)).

(A.1) Lemma:

Für natürliche Zahlen m,n mit $m \leq n$ und monotone Funktionen $g: [m,n] \to \mathbb{R}$, $g \geq 0$ gilt:

$$\int_m^n g(x)dx \leq \Sigma_{i=m}^n g(i) \leq g(m) + \int_m^n g(x)dx + g(n) \quad .$$

Insbesondere gilt für g mit $g(x) = x^\alpha$:

$$\Sigma_{i=1}^n i^\alpha = (\alpha + 1)^{-1} n^{\alpha+1} + O(1 + n^\alpha)$$

für $\alpha \neq -1$ und

$$\Sigma_{i=1}^n i^{-1} = \log n + O(1) \quad .$$

Beweis:

Einfache Überlegungen liefern sofort das Ergebnis:

Für monoton wachsendes g gilt für alle i mit $m < i < n$:

$$\int_{i-1}^{i} g(x)dx \leq g(i) \leq \int_i^{i+1} g(x)dx$$

und

$$\int_{n-1}^{n} g(x)dx \leq g(n) \quad ,$$

woraus durch Aufsummieren über i mit $m < i \leq n$

$$\int_{m}^{n} g(x)dx \leq \Sigma_{i=m+1}^{n} g(i) \leq \int_{m+1}^{n} g(x)dx + g(n)$$

und damit die Behauptung folgt. Für monoton fallendes g gilt Analoges. □

(A.2) Lemma:

Gilt für Funktionen $g,h: [0,\infty) \to \mathbb{R}$ mit $h \geq 0$, h monoton wachsend

und $\lim_{t\to\infty} h(t) = \infty$ und g ist beschränkt auf jedem Kompaktum:

$$g(t) = o(h(t)) \quad ,$$

dann folgt bereits:

$$\sup_{0\leq s\leq t} |g(s)| = o(h(t)) \quad .$$

Beweis:

Nach Voraussetzung existiert für alle $\varepsilon > 0$ ein $T(\varepsilon)$, so daß für

$t \geq T(\varepsilon)$

$$|g(t)| \leq \varepsilon h(t)$$

gilt. Auf Grund der Monotonie von h ergibt sich

$$\sup_{0\leq s\leq t} |g(s)| \leq C(\varepsilon) + \varepsilon h(t)$$

mit $C(\varepsilon) = \sup_{0\leq s\leq T(\varepsilon)} |g(s)| < \infty$. Daraus folgt wegen $\lim_{t\to\infty} h(t)^{-1} = 0$

$$0 \leq \lim_{t\to\infty} \sup (h(t)^{-1} \sup_{0\leq s\leq t} |g(s)|) \leq \varepsilon$$

für alle $\varepsilon > 0$ und damit die Behauptung. □

(A.3) Lemma:

Sind $(x_n)_{n\in\mathbb{N}}$, $(a_n)_{n\in\mathbb{N}}$ Folgen nicht-negativer reeller Zahlen mit $\lim_{n\to\infty} a_n = 0$ und

$$x_{n+1} \leq \max((1-\delta)x_n, a_n)$$

für ein $\delta > 0$, dann gilt

$$\lim_{n\to\infty} x_n = 0 \quad .$$

Beweis:

O.B.d.A. können wir voraussetzen, daß die a_n monoton fallend sind, denn die Voraussetzungen sind mit a_n auch für $a_n' := \sup_{m\geq n} a_n$ an Stelle von a_n erfüllt. Dann gilt für $n > m$ nach Iteration der Voraussetzung:

$$x_n \leq \max((1-\delta)^{n-m}x_m, a_m) \quad .$$

Für alle $\varepsilon > 0$ existiert nun ein $m = m(\varepsilon)$ mit $a_m < \varepsilon$. Daraus folgt:

$$0 \leq \lim_{n\to\infty} \sup x_n \leq x_m \lim_{n\to\infty} \sup(1-\delta)^{n-m} + a_m < \varepsilon$$

für alle $\varepsilon > 0$ und damit die Behauptung. □

II) Einige Ergebnisse aus dem Funsdkreis fast sicherer Invarianzprinzipien:

Im Kapitel 5 hatten wir fast sichere Invarianzprinzipien für die Fehlersummen vorausgesetzt, um konkretere Ergebnisse zu erhalten. In diesem Abschnitt werden wir nun kurze Beweise für die dort benutzten Resultate angeben. Zuerst werden wir eine etwas modifizierte Form von Ruppert's (1982) Lemma 4.1 über ein fast sicheres Invarianzprinzip gewichteter Fehlersummen mit Gewichten n^{α} zeigen. Insbesondere erhalten wir auch eine Aussage für den Fall $\alpha = -\frac{1}{2}$.

(A.4) Satz:

Erfüllen Zufallsvariablen U_n ein fast sicheres Invarianzprinzip: Es existieren $\sigma > 0$ und eine Brown'sche Bewegung B, so daß gilt:

(A) $\Sigma_{i \le t} U_i = \sigma B(t) + o(t^{\frac{1}{2}})$ f.s. ,

dann gilt für $\alpha < -\frac{1}{2}$:

(1) $\Sigma_{i=1}^{\infty} i^{\alpha} U_i$ konvergiert f.s.

und für $\alpha > -\frac{1}{2}$:

Es existiert eine Brown'sche Bewegung B_{α}, so daß gilt:

(2) $\Sigma_{i \le t} i^{\alpha} U_i = \sigma(2\alpha + 1)^{-\frac{1}{2}} B_{\alpha}(t^{2\alpha+1}) + o(t^{\alpha+\frac{1}{2}})$ f.s.

Ist an Stelle von (A) die etwas stärkere Bedingung

(A*) $\Sigma_{i \le t} U_i = \sigma B(t) + o(t^{\frac{1}{2}}(\log t)^{-\frac{1}{2}})$ f.s.

erfüllt, dann gilt für $\alpha = -\frac{1}{2}$:

(3) $\Sigma_{i \le t} i^{-\frac{1}{2}} U_i = \sigma B_{\alpha}(\log t) + o((\log t)^{\frac{1}{2}})$ f.s.

Beweis:

Als erstes führen wir Hilfsgrößen $c_n(\alpha) := n^\alpha - (n-1)^\alpha$ ein, wobei $c_1(\alpha) := 1$ vereinbart sein soll. Wegen $\Sigma_{j=1}^{i} c_j(\alpha) = i^\alpha$ können wir unsere gewichteten Summen teleskopieren und umschreiben:

$$\Sigma_{i=1}^{n} i^\alpha U_i = \Sigma_{i=1}^{n} \Sigma_{j=1}^{i} c_j(\alpha) U_i = \Sigma_{j=1}^{n} c_j(\alpha) \Sigma_{i=j}^{n} U_i \quad .$$

Nach Voraussetzung gilt

$$\Sigma_{i=1}^{n} U_i = \sigma B(n) + R_n$$

mit $R_n = o(n^{\frac{1}{2}})$ bzw. $R_n = o(n^{\frac{1}{2}}(\log n)^{-\frac{1}{2}})$ f.s. Setzen wir noch $R_0 := 0$ und beachten, daß $B(0) = 0$ f.s. gilt, so folgt

$$\Sigma_{i=j}^{n} U_i = \sigma B(n) - \sigma B(j-1) + R_n - R_{j-1} \quad .$$

Wegen $B(n) - B(j-1) = \Sigma_{i=j}^{n} (B(i) - B(i-1))$ folgt durch Umsummieren und Teleskopieren:

$$\Sigma_{j=1}^{n} c_j(\alpha)(B(n) - B(j-1)) = \Sigma_{i=1}^{n} i^\alpha (B(i) - B(i-1))$$

und damit

$$\Sigma_{i=1}^{n} i^\alpha U_i = \sigma \Sigma_{i=1}^{n} i^\alpha (B(i) - B(i-1)) + n^\alpha R_n + \Sigma_{j=1}^{n} c_j(\alpha) R_{j-1} \quad .$$

Nach Voraussetzung gilt nun $n^\alpha R_n = o(n^{\alpha+\frac{1}{2}})$ f.s., so daß wir diesen Term nicht weiter zu betrachten brauchen, sowie $c_n(\alpha) R_{n-1} = o(n^{\alpha-\frac{1}{2}})$ f.s. Daraus folgt für $\alpha < -\frac{1}{2}$ die f.s. Konvergenz von $\Sigma_{j=1}^{\infty} c_j(\alpha) R_{j-1}$ und für $\alpha > -\frac{1}{2}$ mit der Integralabschätzung (A.1) und dem Lemma von Toeplitz:

$$\Sigma_{j=1}^{n} c_j(\alpha) R_{j-1} = o(n^{\alpha+\frac{1}{2}}) \quad \text{f.s.}$$

Unter (A*) gilt $c_n(-\frac{1}{2}) R_{n-1} = o(n^{-1}(\log n)^{-\frac{1}{2}})$ f.s. und mit der Abschätzung (A.1) für g mit $g(x) = x^{-1}(\log x)^{-\frac{1}{2}}$ und dem Lemma von Toeplitz folgt:

$$\Sigma_{j=1}^{n} c_j (-\tfrac{1}{2}) R_{j-1} = o((\log n)^{\frac{1}{2}}) \quad \text{f.s.}$$

Es sei angemerkt, daß wir für $\alpha = -\frac{1}{2}$ unter (A) die Größenordnung $o(\log n)$ f.s. erhalten hätten.

Nun bleibt nur noch $\Sigma_{i=1}^{n} i^{\alpha}(B(i) - B(i-1))$ zu untersuchen:

Für $\alpha < -\frac{1}{2}$ betrachten wir die einzelnen Summanden:

$(n^{\alpha}(B(n) - B(n-1)))_{n \in \mathbb{N}}$ ist eine Folge unabhängiger Zufallsvariablen, jeweils mit Erwartungswert 0 , und es gilt:

$$\Sigma_{n=1}^{\infty} \mathrm{Var}(n^{\alpha}(B(n) - B(n-1))) = \Sigma_{n=1}^{\infty} n^{2\alpha} < \infty \quad .$$

Daraus folgt aber bereits die f.s. Konvergenz von $\Sigma_{n=1}^{\infty} n^{\alpha}(B(n) - B(n-1))$ und damit die Behauptung (1).

Für $\alpha \geq -\frac{1}{2}$ wollen wir eine Brown'sche Bewegung B'_{α} konstruieren, für die gilt:

$$B'_{\alpha}(k_n) = \Sigma_{i=1}^{n} i^{\alpha}(B(i) - B(i-1)) \tag{4}$$

an den Stellen $k_n := \Sigma_{i=1}^{n} i^{2\alpha}$. Dazu definieren wir $B'_{\alpha}(0) := 0$ und

$$B'_{\alpha}(t) := B'_{\alpha}(k_{n-1}) + n^{\alpha}(B((n-1) + n^{-2\alpha}(t - k_{n-1})) - B(n-1))$$

für $k_{n-1} \leq t \leq k_n$, $n \in \mathbb{N}$. Wegen $2\alpha \geq 1$ gilt $k_n \uparrow \infty$ und $B'_{\alpha}(t)$ ist für alle $t \geq 0$ definiert. Es läßt sich relativ einfach nachprüfen, daß B'_{α} Brown'sche Bewegung ist. Für $t = k_n$ gilt wegen $k_n - k_{n-1} = n^{2\alpha}$:

$$B'_{\alpha}(k_n) = B'_{\alpha}(k_{n-1}) + n^{\alpha}(B(n) - B(n-1)) \quad ,$$

und iterieren liefert (4). Bislang haben wir also gezeigt:

Zu $\alpha \geq -\frac{1}{2}$ existiert eine Brown'sche Bewegung B'_{α} , so daß für $\alpha > -\frac{1}{2}$

$$\Sigma_{i \leq t} i^{\alpha} U_i = \sigma B'_{\alpha}(k_{[t]}) + o(t^{\alpha + \frac{1}{2}}) \quad \text{f.s.} \tag{5}$$

bzw. für $\alpha = -\frac{1}{2}$

(6) $\Sigma_{i \le t} i^{-\frac{1}{2}} U_i = \sigma B'_\alpha(k_{[t]}) + o((\log t)^{\frac{1}{2}})$ f.s.

gilt.

Mit der Integralabschätzung (A.1) folgt für $\alpha > -\frac{1}{2}$

$$k_{[t]} = (2\alpha + 1)^{-1} t^{2\alpha+1} + O(1 + t^{2\alpha})$$

bzw. für $\alpha = -\frac{1}{2}$

$$k_{[t]} = \log t + O(1) \quad .$$

Ein tieferliegendes Ergebnis für die Zuwächse einer Brown'schen Bewegung (Satz (A.6)) liefert uns nach geeigneter Variablentransformation ($T(t) := (2\alpha + 1)^{-1} t^{2\alpha+1}$ für $\alpha > -\frac{1}{2}$ bzw. $T(t) := \log t$ für $\alpha = -\frac{1}{2}$):

$$|B'_\alpha(k_{[t]}) - B'_\alpha((2\alpha + 1)^{-1} t^{2\alpha+1})| = O((a_{T(t)} \log T(t))^{\frac{1}{2}}) \quad \text{f.s.}$$

für $\alpha > -\frac{1}{2}$ bzw.

$$|B'_\alpha(k_{[t]}) - B'_\alpha(\log t)| = O((a_{T(t)} \log T(t))^{\frac{1}{2}}) \quad \text{f.s.}$$

für $\alpha = -\frac{1}{2}$, mit $a_T = T^{2\alpha/(2\alpha+1)}$ für $\alpha > 0$ und $a_T = 1$ für $-\frac{1}{2} \le \alpha \le 0$.

Daraus ergibt sich für $\alpha > -\frac{1}{2}$

$$B'_\alpha(k_{[t]}) = B'_\alpha((2\alpha + 1)^{-1} t^{2\alpha+1}) + O((t^\alpha + 1)(\log t)^{\frac{1}{2}}) \quad \text{f.s.}$$

und der letzte Term ist von der Größenordnung $o(t^{\alpha + \frac{1}{2}})$. Da mit B'_α auch B_α mit $B_\alpha(s) := (2\alpha + 1)^{-\frac{1}{2}} B'_\alpha((2\alpha + 1)s)$ Brown'sche Bewegung ist, folgt unter Verwendung von (5) die Behauptung (2). Für $\alpha = -\frac{1}{2}$ erhalten wir

$$B'_\alpha(k_{[t]}) = B'_\alpha(\log t) + O((\log\log t)^{\frac{1}{2}}) \quad \text{f.s.} \quad ,$$

und über (6) ergibt sich Behauptung (3) mit $B_\alpha := B'_\alpha$. □

(A.5) Satz:

Genügt eine zufällige Funktion X mit Werten in $\mathcal{D}[0,\infty)$ der Voraussetzung:

(A) es existiert eine Brown'sche Bewegung B und ein $\alpha > 0$, so daß gilt:

$$X(t) = B(t^{\alpha}) + o(t^{\frac{1}{2}\alpha}) \quad \text{f.s.} \quad ,$$

dann gilt:

$$(n^{-\frac{1}{2}\alpha} X(nt))_{t\in[0,1]} \xrightarrow{\mathcal{D}} (B(t^{\alpha}))_{t\in[0,1]} \quad (\text{in } \mathcal{D}[0,1]) \quad .$$

Beweis:

Mit Lemma (A.2) gilt:

$$\sup_{0\le s\le t} |X(s) - B(s^{\alpha})| = o(t^{\frac{1}{2}\alpha}) \quad \text{f.s.}$$

und somit

$$\sup_{0\le t\le 1} |n^{-\frac{1}{2}\alpha} (X(nt) - B((nt)^{\alpha})| \to 0 \quad \text{f.s.}$$

Nun besitzt aber $(n^{-\frac{1}{2}\alpha} B((nt^{\alpha})))_{t\in[0,1]}$ für alle n dieselbe Verteilung wie $(B(t^{\alpha}))_{t\in[0,1]}$ und die Behauptung folgt mit einem bekannten Ergebnis (vgl. Billingsley (1968), S. 25). □

(A.6) Satz (vgl. Csörgö und Révész (1981), S. 30)

Ist B eine Brown'sche Bewegung und a_T , $T \ge 0$, eine monotone, nichtfallende Funktion von T , die die Voraussetzungen

(A) $0 < a_T \le T$,

(B) Ta_T^{-1} ist monoton nichtfallend

erfüllt, dann gilt

$$\sup_{|s|\le a_T} |B(T + s) - B(T)| = O((a_T(\log \frac{T}{a_T} + \log\log T))^{\frac{1}{2}}) \quad \text{f.s.}$$

III) Einige interessierende Ungleichungen für mischende Folgen von Zufallsvariablen:

Ist (Ω,A,P) der zugrunde liegende Wahrscheinlichkeitsraum und sind $\mathcal{B}$ und $\mathcal{C}$ zwei Unter-σ-Algebren von A, dann definieren wir vier Maße der Abhängigkeit von $\mathcal{B}$ und $\mathcal{C}$ (zu Definition und Eigenschaften vgl. Hall und Heyde (1980), S. 147 f, S. 276 ff, und Ibragimov und Rozanov (1978), S. 108 ff):

$$\alpha(\mathcal{B},\mathcal{C}) := \sup_{B\in\mathcal{B},C\in\mathcal{C}} |P(B \cap C) - P(B)P(C)| \quad ;$$

$$\beta(\mathcal{B},\mathcal{C}) := E(\sup_{C\in\mathcal{C}} |P(C|\mathcal{B}) - P(C)|) \quad ;$$

$$\varphi(\mathcal{B},\mathcal{C}) := \sup_{B\in\mathcal{B},C\in\mathcal{C}:P(B)>0} |P(C|B) - P(C)| \quad ;$$

$$\rho(\mathcal{B},\mathcal{C}) := \sup |E(XY)| \quad ,$$

wobei das Supremum für den maximalen Korrelationskoeffizienten $\rho(\mathcal{B},\mathcal{C})$ über alle $\mathcal{B}$-meßbaren Funktionen X und alle $\mathcal{C}$-meßbaren Funktionen Y mit $E(X) = E(Y) = 0$ und $E(X^2) = E(Y^2) = 1$ gebildet wird. Es gelten die Ungleichungsketten:

$$\alpha(\mathcal{B},\mathcal{C}) \leq \beta(\mathcal{B},\mathcal{C}) \leq \varphi(\mathcal{B},\mathcal{C}) \quad ,$$

$$4\alpha(\mathcal{B},\mathcal{C}) \leq \rho(\mathcal{B},\mathcal{C}) \leq 2\varphi(\mathcal{B},\mathcal{C})^{\frac{1}{2}} \quad ,$$

und zwischen $\beta(\mathcal{B},\mathcal{C})$ und $\rho(\mathcal{B},\mathcal{C})$ ist kein Zusammenhang bekannt. Offensichtlich gilt weiterhin:

$$\alpha(\mathcal{B},\mathcal{C}) = \beta(\mathcal{B},\mathcal{C}) = \varphi(\mathcal{B},\mathcal{C}) = \rho(\mathcal{B},\mathcal{C}) = 0 \quad ,$$

falls $\mathcal{B}$ und $\mathcal{C}$ unabhängig sind. Die Ungleichung $\rho(\mathcal{B},\mathcal{C}) \leq 2\varphi(\mathcal{B},\mathcal{C})^{\frac{1}{2}}$ ergibt sich aus folgendem bekannten Satz mit $p = q = 2$:

(A.7) Satz:

Ist X eine $\mathcal{B}$-meßbare Zufallsvariable mit $E(|X|^p) < \infty$ und Y eine $\mathcal{C}$-meßbare Zufallsvariable mit $E(|X|^q) < \infty$ für $p,q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$, dann gilt:

$$|E(XY) - E(X)E(Y)| \leq 2\varphi(\mathcal{B},\mathcal{C})^{1/p} E(|X|^p)^{1/p} E(|Y|^q)^{1/q} \quad .$$

Den Beweis dieses an die Hölder-Ungleichung erinnernden Ergebnisses findet man z.B. bei Hall und Heyde (1980), S. 278 ff.

Für eine Folge $(U_n)_{n\in\mathbb{N}}$ von Zufallsvariablen definieren wir

$$\alpha(n) := \sup_{m\in\mathbb{N}} \alpha(A_1^m , A_{m+n}^\infty) \quad ,$$

wobei mit A_1^m die von den Zufallsvariablen U_i , $i = 1,\ldots,m$, und mit A_m^∞ die von U_i , $i \geq m$, erzeugten σ-Algebren bezeichnet sein sollen, und wir sagen:

$(U_n)_{n\in\mathbb{N}}$ ist α-mischend (oder: stark mischend), falls $\lim_{n\to\infty} \alpha(n) = 0$ gilt.

Analog definieren wir:

$(U_n)_{n\in\mathbb{N}}$ ist β-mischend (oder: absolut regulär), falls $\lim_{n\to\infty} \beta(n) = 0$ für $\beta(n) := \sup_{m\in\mathbb{N}} \beta(A_1^m , A_{m+n}^\infty)$ gilt;

$(U_n)_{n\in\mathbb{N}}$ ist φ-mischend (oder: gleichmäßig stark mischend), falls $\lim_{n\to\infty} \varphi(n) = 0$ für $\varphi(n) := \sup_{m\in\mathbb{N}} \varphi(A_1^m , A_{m+n}^\infty)$ gilt; und

$(U_n)_{n\in\mathbb{N}}$ ist ρ-mischend (oder: vollständig regulär), falls $\lim_{n\to\infty} \rho(n) = 0$ für $\rho(n) := \sup_{m\in\mathbb{N}} \rho(A_1^m , A_{m+n}^\infty)$ gilt.

Offensichtlich sind unabhängige Folgen φ-mischend, φ-mischende ρ-mischend usw. Andererseits ist die Definition von ρ-mischend allgemein genug, um in der Praxis betrachtete Prozesse, wie z.B. gewisse autoregressive oder "moving average" Prozesse zu umfassen.

Für φ-mischende und ρ-mischende Folgen werden wir unter zusätzlichen Regularitätsbedingungen Abschätzungen der Form

$$E(|\Sigma_{i=a+1}^{a+m} U_i|^p) \leq C m^{\frac{1}{2}p} L(m) \quad ,$$

für alle $a,m \in \mathbb{N}$, angeben, wobei L eine geeignete Funktion mit $L(m) = O(m^{\varepsilon})$ für alle $\varepsilon > 0$ ist. Die von uns benutzte Beweistechnik wurde erstmals von Doob (1953), S. 225, für gewisse Markov-Ketten angewandt und später von Ibragimov (1962) auf stationäre φ-mischende Folgen und ebenfalls von Ibragimov (1975) auf stationäre ρ-mischende Folgen übertragen. Zur Demonstration dieser Technik werden wir quasi als Zugabe eine Marcinkiewicz-Zygmund-Ungleichung für φ-mischende Folgen zeigen und die anderen Beweise auf einige wenige Abschätzungen beschränken. ß-mischende Folgen werden in diesem Abschnitt nicht weiter behandelt; sie sind nur der Vollständigkeit halber aufgeführt, da sie im dritten Teil des Kapitels 8 auftreten. Zum Abschluß werden wir noch eine Abschätzung für stationäre α-mischende Folgen zitieren, die für den Fall, daß p eine gerade natürliche Zahl ist, von Yoshihara (1978) und für beliebiges reelles $p \geq 2$ von Yokoyama (1980) bewiesen wurde.

(A.8) Satz:

Genügt eine φ-mischende Folge $(U_n)_{n\in\mathbb{N}}$ von Zufallsvariablen und ein $p \geq 2$ den Voraussetzungen:

(A) $E(|U_n|^p) < \infty$ für alle n ,

(B) es ex. eine Konstante C_2 und eine nichtfallende Funktion $L: \mathbb{N} \to \mathbb{R}$ mit $L(1) \geq 1$, so daß für alle a und n gilt:
$E(|\Sigma_{i=a+1}^{a+n} U_i|^2) \leq C_2 L(n) \Sigma_{i=a+1}^{a+n} E(U_i^2)$,

dann existiert eine Konstante C_p , die nur von p und den Mischungskoeffizienten $(\varphi(k))_{k\in\mathbb{N}}$ abhängt, so daß für alle a und n gilt:

$$E(|\Sigma_{i=a+1}^{a+n} U_i|^p) \leq C_p L(n)^{\frac{1}{2}p} n^{\frac{1}{2}p-1} \Sigma_{i=a+1}^{a+n} E(|U_i|^p) \quad .$$

Bemerkung:

Es wäre wünschenswert gewesen, unter den Voraussetzungen (A) und (B) eine Abschätzung der Form

$$E(|\Sigma_{i=a+1}^{a+n} U_i|^p) \leq C_p L(n)^{\frac{1}{2}p} E(|\Sigma_{i=a+1}^{a+n} U_i^2|^{\frac{1}{2}p})$$

zu beweisen, aus der dann die Behauptung des Satzes folgt (vgl. Bemerkung i) zu Satz (6.5)). Unsere Beweismethode ist dafür aber nicht geeignet. Wir könnten lediglich

$$E(|\Sigma_{i=a+1}^{a+n} U_i|^p) \leq C_p L(n)^{\frac{1}{2}p} n^{\varepsilon} E(|\Sigma_{i=a+1}^{a+n} U_i^2|^{\frac{1}{2}p})$$

für alle $\varepsilon > 0$ zeigen, was für manche Anwendungen ausreichend sein mag, wollen jedoch nicht auf diesen Beweis näher eingehen.

Beweis:

Nach Voraussetzung (B) ist der Satz für $p = 2$ richtig. Wir werden induktiv zeigen:

Gilt der Satz für eine natürliche Zahl $m \geq 2$, d.h. $p = m$, dann gilt er auch für $p = m + \delta$ mit $0 < \delta \leq 1$.

Da die Voraussetzung (A) mit p auch für $p' \leq p$ erfüllt ist, folgt die Behauptung des Satzes.

Nun führen wir den Beweis des Induktionsschrittes durch:

Als erstes werden wir zeigen:

Für alle $\varepsilon > 0$ existiert eine Konstante c und ein $k \in \mathbb{N}$, so daß für alle $n \geq k$ und alle $a \in \mathbb{N}$ gilt:

$$(1) \quad E(|\Sigma_{i=a+1}^{a+n} U_i + \Sigma_{i=a+n+k+1}^{a+2n} U_i|^{m+\delta})$$

$$\leq (1+\varepsilon)(E(|\Sigma_{i=a+1}^{a+n} U_i|^{m+\delta}) + E(|\Sigma_{i=a+n+k+1}^{a+2n} U_i|^{m+\delta}))$$

$$+ cL(2n)^{\frac{1}{2}(m+\delta)} n^{\frac{1}{2}(m+\delta)-1} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}) \quad .$$

Zur Schreibvereinfachung setzen wir $\alpha := (m+\delta)^{-1}$ und

$$S_1 := \Sigma_{i=a+1}^{a+n} U_i \ , \ S_2 := \Sigma_{i=a+n+k+1}^{a+2n} U_i \ ,$$

$$T_1 := (n^{\frac{1}{2}m-1} \Sigma_{i=a+1}^{a+n} E(|U_i|^m))^{1/m} \ ,$$

$$T_2 := ((n-k)^{\frac{1}{2}m-1} \Sigma_{i=a+n+k+1}^{a+2n} E(|U_i|^m))^{1/m} \quad .$$

Unter Ausnutzung von $|x+y|^\delta \leq |x|^\delta + |y|^\delta$ für $\delta \leq 1$ gilt:

$$(2) \quad E(|S_1 + S_2|^{m+\delta}) \leq E(|S_1 + S_2|^m (|S_1|^\delta + |S_2|^\delta))$$

$$= E(|S_1|^{m+\delta}) + E(|S_2|^{m+\delta})$$

$$+ \Sigma_{j=0}^{m-1} \binom{m}{j} E(|S_1|^{j+\delta} |S_2|^{m-j})$$

$$+ \Sigma_{j=1}^{m} \binom{m}{j} E|S_1|^j |S_2|^{m-j+\delta}) \quad .$$

Nach Satz (A.7) gilt für $0 \leq j < m$:

$$(3) \quad E(|S_1|^{j+\delta} |S_2|^{m-j}) \leq E(|S_1|^{j+\delta}) E(|S_2|^{m-j})$$

$$+ 2\varphi(k)^{(j+\delta)\alpha} E(|S_1|^{m+\delta})^{(j+\delta)\alpha} E(|S_2|^{m+\delta})^{(m-j)\alpha} \quad .$$

Mit Hölder-Ungleichung und Induktionsvoraussetzung folgt:

$$(4) \quad E(|S_1|^{j+\delta}) \leq E(|S_1|^m)^{(j+\delta)/m}$$

$$\leq (C_m L(n)^{\frac{1}{2}m} T_1^m)^{(j+\delta)/m}$$

und analog unter Beachtung von $L(n-k) \leq L(n)$:

$$E(|S_2|^{m-j}) \leq (C_m L(n)^{\frac{1}{2}m} T_2^m)^{(m-j)/m} .$$

Insgesamt ergibt sich somit:

$$(5) \qquad E(|S_1|^{j+\delta}) E(|S_2|^{m-j})$$

$$\leq C_m^{(m+\delta)/m} L(n)^{\frac{1}{2}(m+\delta)} T_1^{j+\delta} T_2^{m-j} .$$

Nun gilt:

$$(6) \qquad \Sigma_{j=0}^{m-1} \binom{m}{j} T_1^{j+\delta} T_2^{m-j} \leq T_1^{\delta} \Sigma_{j=0}^{m} \binom{m}{j} T_1^j T_2^{m-j}$$

$$= T_1^{\delta} (T_1 + T_2)^m$$

$$\leq (T_1 + T_2)^{m+\delta}$$

und analog:

$$(7) \qquad \Sigma_{j=0}^{m-1} \binom{m}{j} E(|S_1|^{m+\delta})^{(j+\delta)\alpha} E(|S_2|^{m+\delta})^{(m-j)\alpha}$$

$$\leq (E(|S_1|^{m+\delta})^{\alpha} + E(|S_2|^{m+\delta})^{\alpha})^{m+\delta}$$

$$\leq 2^{m+\delta} (E(|S_1|^{m+\delta}) + E(|S_2|^{m+\delta})) .$$

Aus der Definition von T_1 und T_2 folgt:

$$(8) \qquad (T_1 + T_2)^m \leq 2^m (T_1^m + T_2^m)$$

$$\leq 2^m n^{\frac{1}{2}m-1} \Sigma_{i=a+1}^{a+2n} E(|U_i|^m) .$$

Die Hölder-Ungleichung für Summen liefert:

$$(9) \qquad \Sigma_{i=a+1}^{a+2n} E(|U_i|^m) \leq \Sigma_{i=a+1}^{a+2n} (E(|U_i|^{m+\delta}))^{m\alpha}$$

$$\leq (2n)^{\delta\alpha} (\Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}))^{m\alpha} ,$$

so daß sich aus (8) und (9) mit $(\frac{1}{2}m - 1)(m + \delta) + \delta = (\frac{1}{2}(m + \delta) - 1)m$:

(10) $$(T_1 + T_2)^{m+\delta} \le 2^{m+\delta} n^{(\frac{1}{2}m-1)(m+\delta)/m} (2n)^{\delta/m} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta})$$
$$\le 2^{m+2\delta} n^{\frac{1}{2}(m+\delta)-1} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta})$$

ergibt. Wegen $\varphi(k) \le 1$ gilt $\varphi(k)^{(j+\delta)\alpha} \le \varphi(k)^{\delta\alpha}$, und wir erhalten

(11) $$\Sigma_{j=0}^{m-1} E(|S_1|^{j+\delta} |S_2|^{m-j})$$
$$\le \frac{1}{2} cL(n)^{\frac{1}{2}(m+\delta)} n^{\frac{1}{2}(m+\delta)-1} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta})$$
$$+ 2^{m+\delta+1} \varphi(k)^{\delta\alpha} (E(|S_1|^{m+\delta}) + E(|S_2|^{m+\delta}))$$

mit $\frac{1}{2} c = 2^{m+2\delta} C_m^{(m+\delta)/m}$. Die zweite Summe in (2) läßt sich vollkommen analog abschätzen, und wir erhalten:

(12) $$E(|S_1 + S_2|^{m+\delta})$$
$$\le (1 + 2^{m+\delta+2} \varphi(k)^{\delta\alpha}) (E(|S_1|^{m+\delta}) + E(|S_2|^{m+\delta}))$$
$$+ cL(n)^{\frac{1}{2}(m+\delta)} n^{\frac{1}{2}(m+\delta)-1} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}) \quad .$$

Nun gilt $L(n) \le L(2n)$, und da die Folge φ-mischend ist, existiert ein k, so daß $2^{m+\delta+2} \varphi(k)^{\delta\alpha} < \varepsilon$ gilt, womit die Behauptung (1) bewiesen ist.

Als nächstes werden wir zeigen:

Für alle $\varepsilon' > 0$ existiert eine Konstante c', so daß für alle $a,n \in \mathbb{N}$ gilt:

(13) $$E(|\Sigma_{i=a+1}^{a+2n} U_i|^{m+\delta})$$
$$\le (1 + \varepsilon') (E(|\Sigma_{i=a+1}^{a+n} U_i|^{m+\delta}) + E(|\Sigma_{i=a+n+1}^{a+2n} U_i|^{m+\delta})$$
$$+ c' L(2n)^{\frac{1}{2}(m+\delta)} n^{\frac{1}{2}(m+\delta)-1} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}) \quad .$$

Für alle $n \ge k$ gilt mit der Minkowski-Ungleichung:

(14) $$E(|\Sigma_{i=a+1}^{a+2n} U_i|^{m+\delta})$$

$$\leq (E(|\Sigma_{i=a+1}^{a+n} U_i + \Sigma_{i=a+n+k+1}^{a+2n} U_i|^{m+\delta})^{\alpha} + E(|\Sigma_{i=a+n+1}^{a+n+k} U_i|^{m+\delta})^{\alpha})^{m+\delta}$$

wobei wieder $\alpha = (m + \delta)^{-1}$ sein soll. Nun gilt weiterhin:

(15) $$E(|\Sigma_{i=a+n+1}^{a+n+k} U_i|^{m+\delta}) \leq k^{m+\delta} \Sigma_{i=a+n+1}^{a+n+k} E(|U_i|^{m+\delta}) \quad .$$

Wegen $|x + y|^{m+\delta} \leq (1 + \varepsilon)^{m+\delta}|x|^{m+\delta} + (1 + \varepsilon^{-1})^{m+\delta}|y|^{m+\delta}$ für alle $\varepsilon > 0$ erhalten wir aus (14) und (15)

(16) $$E(|\Sigma_{i=a+1}^{a+2n} U_i|^{m+\delta})$$

$$\leq (1 + \varepsilon)^{m+\delta} E(|\Sigma_{i=a+1}^{a+n} U_i + \Sigma_{i=a+n+k+1}^{a+2n} U_i|^{m+\delta})$$

$$+ (1 + \varepsilon^{-1})^{m+\delta} k^{m+\delta} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}) \quad .$$

Mit derselben Argumentation folgt

(17) $$E(|\Sigma_{i=a+n+k+1}^{a+2n} U_i|^{m+\delta})$$

$$\leq (1 + \varepsilon)^{m+\delta} E(|\Sigma_{i=a+n+1}^{a+2n} U_i|^{m+\delta})$$

$$+ (1 + \varepsilon^{-1})^{m+\delta} k^{m+\delta} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}) \quad .$$

Fassen wir nun (16), (3) und (17) zusammen, dann ist für alle $\varepsilon > 0$ die Existenz eines $k \in \mathbb{N}$ und eines c' gesichert, so daß für alle $n \geq k$ und alle a gilt:

(18) $$E(|\Sigma_{i=a+1}^{a+2n} U_i|^{m+\delta})$$

$$\leq (1 + \varepsilon)^{2(m+\delta)+1} (E(|\Sigma_{i=a+1}^{a+n} U_i|^{m+\delta}) + E(|\Sigma_{i=a+n+1}^{a+2n} U_i|^{m+\delta}))$$

$$+ c' L(2n)^{\frac{1}{2}(m+\delta)} n^{\frac{1}{2}(m+\delta)-1} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}) \quad .$$

Dabei ist $c' := (1 + \varepsilon^{-1})^{m+\delta} k^{m+\delta} + (1 + \varepsilon)^{m+\delta} c + (1 + \varepsilon)^{m+\delta+1} (1 + \varepsilon^{-1})^{m+\delta} k$ mit c und k aus (1). Weiterhin gilt für $n < k$ analog zu (15):

$$E(|\Sigma_{i=a+1}^{a+2n} U_i|^{m+\delta}) \leq (2k)^{m+\delta} \Sigma_{i=a+1}^{a+2n} E(|U_i|^{m+\delta}) \quad ,$$

und wegen $2 \leq (1+\varepsilon)(1+\varepsilon^{-1})$ ist (18) mit c' wie oben definiert auch für $n < k$ richtig. Nun gibt es zu $\varepsilon' > 0$ offensichtlich $\varepsilon > 0$ mit $(1+\varepsilon)^{2(m+\delta)+1} \leq (1+\varepsilon')$, und Behauptung (13) folgt.

Wir erhalten aus (13) für $n = 2^r$:

$$(19) \qquad E(|\Sigma_{i=a+1}^{a+2^r} U_i|^{m+\delta})$$

$$\leq (1+\varepsilon')(E|\Sigma_{i=a+1}^{a+2^{r-1}} U_i|^{m+\delta}) + E(|\Sigma_{i=a+2^{r-1}+1}^{a+2^r} U_i|^{m+\delta}))$$

$$+ c'L(2^r)^{\frac{1}{2}(m+\delta)} 2^{r(\frac{1}{2}(m+\delta)-1)} \Sigma_{i=a+1}^{a+2^r} E(|U_i|^{m+\delta})$$

und induktiv:

$$(20) \qquad E(|\Sigma_{i=a+1}^{a+2^r} U_i|^{m+\delta})$$

$$\leq (1+\varepsilon')^r \Sigma_{i=a+1}^{a+2^r} E(|U_i|^{m+\delta})$$

$$+ \Sigma_{j=0}^{r-1} (1+\varepsilon')^j c'L(2^{r-j})^{\frac{1}{2}(m+\delta)} 2^{(r-j)(\frac{1}{2}(m+\delta)-1)} \Sigma_{i=a+1}^{a+2^r} E(|U_i|^{m+\delta})$$

$$\leq c'2(2^r)^{\frac{1}{2}(m+\delta)} 2^{r(\frac{1}{2}(m+\delta)-1)} \Sigma_{i=a+1}^{a+2^r} E|U_i|^{m+\delta})$$

$$\cdot \Sigma_{j=0}^{r} (1+\varepsilon')^j 2^{-j(\frac{1}{2}(m+\delta)-1)} .$$

Wegen $m + \delta > 2$ existiert ein $\varepsilon' > 0$, so daß $1 + \varepsilon' < 2^{\frac{1}{2}(m+\delta)-1}$ gilt. Folglich gilt für den letzten Term in (19):

$$\Sigma_{j=0}^{r} (1+\varepsilon')^j 2^{-j(\frac{1}{2}(m+\delta)-1)} \leq c^* < \infty$$

mit $c^* = (1 - (1+\varepsilon')2^{-\frac{1}{2}(m+\delta)+1})^{-1}$. Damit ist der Induktionsschritt für $n = 2^r$ bewiesen, und die Konstante $c'c^*$ hängt nur von $m + \delta$ und den Mischungskoeffizienten $(\varphi(k))_{k\in\mathbb{N}}$ ab. Für allgemeines n mit $2^r \leq n < 2^{r+1}$ betrachten wir die Binärdarstellung $n = \Sigma_{j=0}^{r} \alpha_j 2^{r-j}$ mit $\alpha_j \in \{0,1\}$. Sind $\nu(k) = \Sigma_{j=0}^{k-1} \alpha_j 2^{r-j}$, $0 \leq k \leq r+1$, die Partialsummen der Binärdarstellung, dann gilt:

$$\Sigma_{i=a+1}^{a+n} U_i = \Sigma_{k=0}^{r} \Sigma_{a+\nu(k)+1}^{a+\nu(k+1)} U_i$$

und $\nu(k+1) - \nu(k) = \alpha_k 2^{r-k}$. Mit der Minkowski-Ungleichung folgt nun (mit $\alpha := (m+\delta)^{-1}$):

$$E(|\Sigma_{i=a+1}^{a+n} U_i|^{m+\delta})$$

$$\leq (\Sigma_{k=0}^{r} E(|\Sigma_{a+\nu(k)+1}^{a+\nu(k+1)} U_i|^{m+\delta})^{\alpha})^{m+\delta} \quad ,$$

und wir können (20) auf die Teilsummen anwenden, falls $\alpha_k = 1$ gilt:

$$E(|\Sigma_{a+\nu(k)+1}^{a+\nu(k+1)} U_i|^{m+\delta})$$

$$\leq c'c^* L(2^{r-k})^{\frac{1}{2}(m+\delta)} 2^{(r-k)(\frac{1}{2}(m+\delta)-1)} \Sigma_{i=a+\nu(k)+1}^{a+\nu(k+1)} E(|U_i|^{m+\delta}) \quad ,$$

wobei diese Ungleichung für $\alpha_k = 0$ trivialerweise richtig ist.

Einsetzen liefert:

(21) $$E(|\Sigma_{i=a+1}^{a+n} U_i|^{m+\delta})$$

$$\leq c'c^* L(n) 2^{r(\frac{1}{2}(m+\delta)-1)} (\Sigma_{k=0}^{r} 2^{-k(\frac{1}{2}-\alpha)})^{m+\delta} \Sigma_{i=a+1}^{a+n} E(|U_i|^{m+\delta}) \quad .$$

Wegen $\alpha < \frac{1}{2}$ gilt

$$\Sigma_{k=0}^{r} 2^{-k(\frac{1}{2}-\alpha)} < (1 - 2^{-\frac{1}{2}+\alpha})^{-1} < \infty \quad ,$$

und aus (21) folgt die Behauptung mit $C_{m+\delta} = c'c^*(1 - 2^{-\frac{1}{2}+\alpha})^{-1}$. □

Mit derselben Methode läßt sich für φ-mischende Folgen unter etwas abgeänderten Voraussetzungen eine Abschätzung der Form $\sup_{a \in \mathbb{N}} E(|\Sigma_{i=a+1}^{a+n} U_i|^p) \leq C_p n^{\frac{1}{2}p} L(n)^{\frac{1}{2}p}$ beweisen. Wir werden diese Aussage aber sogar für ρ-mischende Folgen formulieren:

(A.9) Satz:

Genügt eine ρ-mischende Folge $(U_n)_{n \in \mathbb{N}}$ von Zufallsvariablen und ein $p \geq 2$ den Voraussetzungen

(A) $\sup_{n\in\mathbb{N}} E(|U_n|^p) < \infty$,

(B) es ex. eine Konstante C_2 und eine nichtfallende Funktion $L: \mathbb{N} \to \mathbb{R}$ mit $L(1) \geq 1$, so daß für alle n gilt:

$$\sup_{a\in\mathbb{N}} E(|\Sigma_{i=a+1}^{a+n} U_i|^2) \leq C_2 n L(n) \quad ,$$

dann existiert eine Konstante C_p , die nur von p und den Mischungskoeffizienten $(\rho(k))_{k\in\mathbb{N}}$ abhängt, so daß für alle n gilt:

$$a(n) := \sup_{a\in\mathbb{N}} E(|\Sigma_{i=a+1}^{a+n} U_i|^p) \leq C_p n^{\frac{1}{2}p} L(n)^{\frac{1}{2}p} \quad .$$

Bemerkung:

Ist $(U_n)_{n\in\mathbb{N}}$ stationär und existiert $E(|U_n|^p) < \infty$, dann ist (A) trivialerweise und (B) nach Ibragimov (1975) sogar mit einer langsam variierenden Funktion erfüllt (vgl. auch Hall und Heyde (1980), S. 148). Dazu muß man beachten, daß es zu einer langsam variierenden Funktion $h: [0,\infty) \to \mathbb{R}$ immer eine nichtfallende langsam variierende Funktion $L: [0,\infty) \to [1,\infty)$ mit $h(x) = O(L(x))$ gibt. Dieses folgt aber aus einem bekannten Kriterium für langsame Variation (vgl. Loève (1977) I, S. 357). Ist andererseits L langsam variierend, dann gilt $L(n)^{\frac{1}{2}p} = O(n^{\varepsilon})$ für alle $\varepsilon > 0$, und wir erhalten eine Abschätzung, wie wir sie im Kapitel 6 benötigten.

Beweis:

Die Idee ist dieselbe wie im Beweis von Satz (A.8). Da wir jedoch für ρ-mischende Folgen nicht die Aussage des Satzes (A.7) zur Verfügung haben, können wir an Stelle von (A.8) (1) lediglich zeigen:

Für alle $\varepsilon > 0$ existiert eine Konstante c und ein $k \in \mathbb{N}$, so daß für alle $n \geq k$ und alle a gilt (mit $\alpha := (m+\delta)^{-1}$):

$$(1)\qquad E(|\Sigma_{i=a+1}^{a+n} U_i + \Sigma_{i=a+n+k+1}^{a+2n+k} U_i|^{m+\delta})$$

$$\leq (1+\varepsilon)(E(|\Sigma_{i=a+1}^{a+n} U_i|^{m+\delta}) + E(|\Sigma_{i=a+n+k+1}^{a+2n+k} U_i|^{m+\delta}))$$

$$+ c(nL(n))^{\delta} \max((nL(n))^{\frac{1}{2}(m-\delta)} a(n)^{(m-\delta)\alpha}) \quad .$$

Der Nachweis wird wieder über (A.8) (2) geführt, indem wir $S_1 := \Sigma_{i=a+1}^{a+n} U_i$ und $S_2 := \Sigma_{i=a+n+k+1}^{a+2n+k} U_i$ setzen, jedoch ist die Abschätzung von $E(|S_1|^{j+\delta}|S_2|^{m-j})$ für $0 \leq j < m$ komplizierter, was wir hier für $j + \delta \leq m - j$ exemplarisch andeuten:

Mit der Hölder-Ungleichung gilt:

$$(2)\qquad E(|S_1|^{j+\delta}|S_2|^{m-j}) = E(|S_1 S_2|^{j+\delta}|S_2|^{m-2j-\delta})$$

$$\leq E(|S_1 S_2|^{\frac{1}{2}(m+\delta)})^{2(j+\delta)\alpha} E(|S_2|^{m+\delta})^{(m-2j-\delta)\alpha} \quad ,$$

und, da die Folge ρ-mischend ist, gilt weiterhin:

$$(3)\qquad E(|S_1 S_2|^{\frac{1}{2}(m+\delta)}) \leq E(|S_1|^{\frac{1}{2}(m+\delta)}) E(|S_2|^{\frac{1}{2}(m+\delta)})$$

$$+ \rho(k) E(|S_1|^{m+\delta})^{\frac{1}{2}} E(|S_2|^{m+\delta})^{\frac{1}{2}} \quad .$$

Unter Ausnutzung der Induktionsvoraussetzung:

$$E(|S_i|^{\frac{1}{2}(m+\delta)}) \leq E(|S_i|^m)^{\frac{1}{2}(m+\delta)/m}$$

$$\leq C_m^{\frac{1}{2}(m+\delta)/m} (nL(n))^{\frac{1}{2}(m+\delta)} \quad ,$$

$i = 1,2$, folgt aus (2) und (3):

$$(4)\qquad E(|S_1|^{j+\delta}|S_2|^{m-j})$$

$$\leq C_m^{2(j+\delta)/m} (nL(n))^{j+\delta} a(n)^{(m-2j-\delta)\alpha}$$

$$+ \rho(k)^{2(j+\delta)\alpha} E(|S_1|^{m+\delta})^{(j+\delta)\alpha} E(|S_2|^{m+\delta})^{(m-j)\alpha} \quad ,$$

und somit:

$$(5)\qquad E(|S_1|^{j+\delta}|S_2|^{m-j}) \le C_m^{(m+\delta)/m}(nL(n))^{\delta}\max((nL(n))^{\frac{1}{2}(m-\delta)}, a(n)^{(m-\delta)\alpha}) + \rho(k)^{2\delta\alpha}E(|S_1|^{m+\delta})^{(j+\delta)\alpha}E(|S_2|^{m+\delta})^{(m-j)\alpha} .$$

Dieselbe Abschätzung (5) ergibt sich aber auch für $j + \delta > m - j$, und der Rest der Behauptung (1) folgt wie in (A.8).

Analog zu (A.8) (13) erhalten wir aus (1):

Für alle $\varepsilon' > 0$ existiert eine Konstante c', so daß für alle n

$$(6)\qquad a(2n) \le (1 + \varepsilon')2a(n) + c'(nL(n))^{\delta}\max((nL(n)^{\frac{1}{2}(m-\delta)}, a(n)^{(m-\delta)\alpha})$$

gilt. Setzen wir nun $b(n) := \max((nL(n))^{-\frac{1}{2}(m+\delta)}a(n), 1)$, dann folgt aus (6):

$$(7)\qquad b(2n) \le 2(1 + \varepsilon')2^{-\frac{1}{2}(m+\delta)}b(n) + c''b(n)^{(m-\delta)\alpha}$$

mit $c'' \ge \max(c', 1)$. Nun gibt es $\varepsilon' > 0$, so daß $\beta := 2(1 + \varepsilon')2^{-\frac{1}{2}(m+\delta)} < 1$ gilt. Für $n = 2^r$ erhalten wir somit die Abschätzung:

$$(8)\qquad 0 \le b(2^{r+1}) \le \beta b(2^r) + c''b(2^r)^{(m-\delta)\alpha}$$

und daraus ergibt sich, daß die Folge $(b(2^r))_{r\in\mathbb{N}}$ beschränkt ist, und zwar durch $c^* := \max(b(1), ((1 - \beta)^{-1}c'')^{\frac{1}{2}(m+\delta)/\delta})$. Die Behauptung des Satzes gilt also mit der Konstante c^* für $n = 2^r$, und für beliebiges n ergibt sie sich wie in (A.8) über die Binärdarstellung mit $C_{m+\delta} = c^*(1 - 2^{-\frac{1}{2}+\alpha})^{-1}$. □

Die Frage, ob sich auch für ρ-mischende Folgen eine Marcinkiewicz-Zygmund-Ungleichung der im Satz (A.8) beschriebenen Form beweisen läßt, müssen wir ebenso wie die nach einer allgemeinen Abschätzung für α-mischende Folgen in diesem Zusammenhang unbeantwortet lassen. Für den Fall, daß die α-mischende Folge stationär ist, läßt sich unter zusätzlichen Bedingungen sowohl an die Momente, als auch die Mischungskoeffizienten noch folgendes Ergebnis angeben (vgl. Yokoyama (1980)):

(A.10) Satz:

Genügt eine stationär α-mischende Folge $(U_n)_{n\in\mathbb{N}}$ von Zufallsvariablen mit $E(U_n) = 0$, sowie $p \geq 2$ und $\delta > 0$ den Voraussetzungen

(A) $E(|U_n|^{p+\delta}) < \infty$,

(B) $\Sigma_{i=0}^{\infty}(i+1)^{\frac{1}{2}p-1}(\alpha(i))^{\delta/(p+\delta)} < \infty$,

dann existiert eine Konstante C, so daß für alle n gilt:

$$E(|\Sigma_{i=1}^{n}U_i|^p) \leq Cn^{\frac{1}{2}p} \quad .$$

Literatur

Berkes, I. und Philipp, W. (1979):

Approximation theorems for independent and weakly dependent random vectors.
Ann.Probab. 7, 29 - 54.

Billingsley, P. (1968):

Convergence of probability measures.
New York: Wiley.

Blum, J.R. (1954 a):

Approximation methods which converge with probability one.
Ann.Math.Stat. 25, 382 - 386.

Blum, J.R. (1954 b):

Multidimensional stochastic approximation methods.
Ann.Math.Stat. 25, 737 - 744.

Burkholder, D.I. (1956):

On a class of stochastic approximation processes.
Ann.Math.Stat. 27, 1044 - 1059.

Chung, K.L. (1954):

On a stochastic approximation method.
Ann.Math.Stat. 25, 463 - 483.

Clark, D.S. (1984):

Necessary and sufficient conditions for the Robbins-Monro method.
Stochastic Processes Appl. 17, 359 - 367.

Csörgö, M. und Révész, P. (1981):

Strong approximation in probability and statistics.
New York: Academic Press.

Derman, C. and Sacks, J. (1959):

On Dvoretzky's stochastic approximation theorem.
Ann.Math.Stat. 30, 601 - 606.

Doob, J.L. (1953):

Stochastic processes.
New York: Wiley.

Dupac, V. (1966):

Stochastic approximation in the presence of trend.
Czech.Math.J., 16(91), 454 - 461.

Fabian, V. (1968):

On asymptotic normality in stochastic approximation.
Ann.Math.Stat. 39, 1327 - 1332.

Fabian, V. (1971):

Stochastic approximation.
In: Rustagi, J.S.: Optimizing methods in statistics, 439 - 470.
New York: Academic Press.

Fabian, V. (1973)

Asymptotically efficient stochastic approximation;
the RM case.
Ann.Stat. 1, 486 - 495.

Gaposhkin, V.F. und Krasulina, T.P. (1974):

On the law of the iterated logarithm in stochastic
approximation processes.
Theory Probab.Appl. 19, 844 - 850.

Hall, P. und Heyde, C.C. (1980):

Martingale limit theory and its applications.
New York: Academic Press.

Herrndorf, N. (1984):

A functional central limit theorem for ρ-mixing sequences.
J.Multivariate Anal. 15, 141 - 146.

Heyde, C.C. (1974):

On martingale limit theory and strong convergence results for stochastic approximation procedures.
Stochastic Processes Appl. 2, 359 - 370.

Heyde, C.C. (1981):

Invariance principles in statistics.
Int.Stat.Rev. 49, 143 - 152.

Holst, U. (1984):

Convergence of a recursive robust algorithm with strongly regular observations.
Stochastic Processes Appl. 16, 305 - 320.

Huber, P.J. (1981):

Robust statistics.
New York: Wiley.

Ibragimov, I.A. (1962):

Some limit theorems for stationary processes.
Theory Probab.Appl. 7, 349 - 382.

Ibragimov, I.A. (1975):

A note on the central limit theorem for dependent random variables.
Theory Probab.Appl. 20, 135 - 141.

Ibragimov, I.A. und Rozanov, Y.A. (1978):

Gaussian random processes.
New York: Springer.

Kersting, G. (1977):

Almost sure approximation of the Robbins-Monro process by sums of independent random variables. Ann.Probab. 5, 954 - 965.

Kiefer, J. und Wolfowitz, J. (1952):

Stochastic estimation of the maximum of a regression function. Ann.Math.Stat. 23, 462 - 466.

Ljung, L. (1978):

Strong convergence of a stochastic approximation algorithm. Ann.Stat. 6, 680 - 696.

Loève, M. (1977/1978):

Probability Theory, I/II. 4th ed. New York: Springer.

Major, P. (1973):

A law of the iterated logarithm for the Robbins-Monro method. Stud.Sci.Math.Hung. 8, 95 - 102.

Major, P. und Révész, P. (1973):

A limit theorem for the Robbins-Monro approximation. Z.Wahrscheinlichkeitstheor.Verw.Geb. 27, 79 - 86.

Mangoldt, H. v. und Knopp, K. (1974):

Einführung in die Höhere Mathematik. 2. Band, 14. Aufl. Stuttgart: Hirzel.

Martin, R.D. und Masreliez, C.J. (1975):

Robust estimation via stochastic approximation. IEEE Trans.Inf.Theory 21, 263 - 271.

McLeish, D.L. (1976):

Functional and random central limit theorems for the Robbins-Monro process.
J.Appl.Probab. 13, 148 - 154.

Nevel'son, M.B. und Has'minskii, R.Z. (1973):

Stochastic approximation and recursive estimation.
Translation of Mathematical Monographs 47.
Providence (R.I.): Am.Math.Soc.

Nevel'son, M.B. und Khas'minskii, R.Z. (1973/1974):

An adaptive Robbins-Monro procedure.
Autom.Remote Control 34, 1594 - 1607.

Peligrad, M. (1982):

Invariance principles for mixing sequences of random variables.
Ann.Probab. 10, 968 - 981.

Philipp, W. und Stout, W. (1975):

Almost sure invariance principles for partial sums of weakly dependent random variables.
Mem.Am.Math.Soc. 161.

Robbins, H. und Monro, S. (1951):

A stochastic approximation method.
Ann.Math.Stat. 22, 400 - 407.

Ruppert, D. (1982):

Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise.
Ann.Probab. 10, 178 - 187.

Sacks, J. (1958):

Asymptotic distribution of stochastic approximation procedures.
Ann.Math.Stat. 29, 373 - 405.

Schmetterer, L. (1979):

From stochastic approximation to the stochastic theory of optimization.
Ber.Math.-Stat.Sekt.Forschungszent. Graz, 127.

Serfling, R.J. (1970):

Moment inequalities for the maximum cumulative sum.
Ann.Math.Stat. 41, 1227 - 1234.

Stout, W.F. (1974):

Almost sure convergence.
New York: Academic Press.

Strassen, V. (1964):

An invariance principle for the law of the iterated logarithm.
Z.Wahrscheinlichkeitstheor.Verw.Geb. 3, 211 - 226.

Strassen, V. (1966):

Almost sure behavior of independent random variables and martingales.
Proc. 5th Berkeley Symp.Math.Stat.Probab. 1965, 2, 315 - 343.

Stummel, F. und Hainer, K. (1982):

Praktische Mathematik. 2. Aufl.
Stuttgart: Teubner.

Venter, J.H. (1967):

An extension of the Robbins-Monro procedure.
Ann.Math.Stat. 38, 181 - 190.

Walk, H. (1977):

An invariance principle for the Robbins-Monro process in a Hilbert space.
Z.Wahrscheinlichkeitstheor.Verw.Geb. 39, 135 - 150.

Wasan, M.T. (1969):

Stochastic approximation.
Cambridge Tracts in Mathematics and Mathematical Physics 58.
Cambridge: University Press.

Wetherill, G.B. (1975):

Sequential methods in statistics. 2nd ed.
London: Chapman and Hall.

Yokoyama, R. (1980):

Moment bounds for stationary mixing sequences.
Z.Wahrscheinlichkeitstheor.Verw.Geb. 52, 45 - 57.

Yoshihara, K. (1978):

Moment inequalities for mixing sequences.
Kodai Math.J. 1, 316 - 328.

Lebenslauf

Am 26. November 1954 wurde ich als zweiter Sohn des Diplom-Chemikers Dr. Walter Schwabe und der Hausfrau Paula Schwabe, geb. Neumann, in Mainz/Rhein geboren. Dort besuchte ich auch von 1961 bis 1965 die katholische Bekenntnisschule Fürstenberger Hof. Nach einem Umzug ins Westfälische besuchte ich ab 1965 das neusprachliche Pestalozzi-(I)-Gymnasium für Jungen in Unna/Westf. und ab 1968 das mathematisch-naturwissenschaftliche Freiherr-vom-Stein-Gymnasium für Jungen in Hamm/Westf., an welchem ich im Sommer 1973 mein Abitur ablegte.

In der Zeit von 1972 bis 1977 war ich nebenamtlich im Bereich der internationalen und politischen Jugendbildung im Rahmen des deutschen paritätischen Wohlfahrtsverbandes tätig. Zum Wintersemester 1973/74 nahm ich an der Philipps-Universität Marburg das Studium der Mathematik und Soziologie auf und wechselte zum Wintersemester 1974/75 an die Freie Universität Berlin, um Mathematik mit Nebenfach Angewandte Statistik (Betriebswirtschaftslehre) zu studieren. Während meines Studiums arbeitete ich unter anderem als Werkstudent in der Datenverarbeitung bei Schering und als studentischer Mitarbeiter beim Zentralblatt für Mathematik. Im Verlauf meiner dreijährigen Tutorentätigkeit konnte ich erste Lehrerfahrungen sammeln.

Am 28. April 1982 schloß ich das Studium der Mathematik mit dem Diplom ab.

Da ich aufgrund der allgemeinen Stellenpolitik etwas Zeit hatte, konnte ich mich in der Zeit bis Frühjahr 1983 intensiv um meinen am 4. September 1980 geborenen Sohn, Jonas Rothe, kümmern und an Aufbau und Organisation einer Eigeninitiativ-Kindertagesstätte teilnehmen.

Seit Sommersemester 1983 bin ich als wissenschaftlicher Mitarbeiter am III. Mathematischen Institut der Freien Universität Berlin in der Fachrichtung Mathematische Statistik und Wahrscheinlichkeitstheorie beschäftigt.